



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

문학석사 학위논문

**Effects of Within-task Planning Conditions on  
Korean EFL Test-takers' Performance on  
Argumentative Writing Assessment Tasks**

과제 내 계획하기 조건들이 한국인 EFL  
수험자들의 논증적 글쓰기 과제 수행에  
미치는 영향

2013 년 8 월

서울대학교 대학원  
영어영문학과 어학전공  
박정현

**Effects of Within-task Planning Conditions on  
Korean EFL Test-takers' Performance on  
Argumentative Writing Assessment Tasks**

지도 교수 이용원

이 논문을 문학석사 학위논문으로 제출함  
2013년 8월

서울대학교 대학원  
영어영문학과 어학 전공  
박정현

박정현의 문학석사 학위论문을 인준함  
2013년 8월

위 원 장 \_\_\_\_\_ (인)

부위원장 \_\_\_\_\_ (인)

위 원 \_\_\_\_\_ (인)

# **Abstract**

Planning has been considered an essential component of cognitive processes in writing as writers generate and organize ideas, and set goals. Following this reasoning, a number of studies attempted to investigate effects of planning and ended up reporting mixed results. A great deal of testing literature discusses how the task that is involved and the way the task implemented influence test-taker performance on tasks. In this regard, task conditions under which tasks are conducted should be significant consideration. However, there have been relatively few studies conducted on effects of planning conditions on writing in a testing context.

This study aimed to examine the effects of within-task planning conditions on Korean EFL test-takers' performance on argumentative writing assessment tasks. Within-task planning refers to on-line planning activities writers engage in while they are performing a writing task. Twenty-eight Korean university students were divided into two major proficiency groups (i.e., advanced and intermediate) and were asked to perform two argumentative writing assessment tasks under three different planning conditions: no planning; pressured within-task planning; and unpressured within-task planning.

A series of repeated measure analyses of variance (ANOVA) and analyses of covariance (ANCOVA) were conducted. First of all, the analyses of variance examined the statistical significance of differences in writing scores among three different planning conditions using analytic scores and composite scores (i.e., aggregated scores of the six analytic criteria, namely, development of ideas, organization, vocabulary, sentence variety and construction, grammar and usage, and mechanics). Second, the analyses of covariance examined whether test-takers' proficiency was a factor that mediated effects of planning conditions on writing scores. Pre-test scores were used as covariate. Finally, planning patterns and behaviors of Korean EFL test-takers were examined through a pre-test questionnaire, a post-test questionnaire, and interviews. Test-takers' planning sheets were also examined.

The analysis of results indicated that within-task planning conditions had no significant effects on the composite scores as well as analytic writing scores of Korean EFL test-takers. Furthermore, the test-takers' proficiency levels turned out to have no mediating influence on effects of planning conditions on writing scores. In terms of planning behavior, seventy-five percent of the participants responded that planning is important and helpful in organizing key concepts and ideas. Also, it was shown that advanced writers spend more time on planning than intermediate writers.

Based on the findings of this study, one major implication is the assignment of within-task planning time. Since there were no significant effects of three experimental conditions, the present study provided empirical evidence in support of the current writing test format which limits time incorporating all process of writing instead of giving no within-task planning time separately.

**Keywords :** pressured and unpressured within-task planning conditions, effects of planning conditions, argumentative writing tasks, test-takers' proficiency, planning behavior

**Student number:** 2009-22753

# Table of Contents

Abstract .....	i
Tables of Contents .....	ii
List of Tables .....	iv
List of Figures .....	v
 Chapter I . Introduction .....	 1
1.1 Background and motivation .....	1
1.2 Research questions .....	5
1.3 Organization of the thesis.....	5
 Chapter II . Literature Review .....	 7
2.1 A theoretical model of L2 Writing.....	7
2.2 The frameworks of L2 writing assessment.....	9
2.2.1 General framework.....	11
2.2.2 Time limit .....	12
2.2.3 Types of planning .....	16
2.2.4 Planning patterns and behaviors.....	17
2.3 Previous studies on effects of planning in an assessment context.....	20
2.3.1 Speaking .....	20
2.3.2 Writing.....	21
 Chapter III. Method.....	 24
3.1 Participants.....	24
3.2 Raters .....	27
3.3 Examiners.....	28
3.4 Instrument .....	28
3.4.1 Placement test.....	29
3.4.2 Writing tests .....	30
3.4.3 Planning conditions .....	30
3.4.3.1 No planning .....	31
3.4.3.2 Pressured within-task planning .....	31
3.4.3.3 Unpressured within-task planning.....	32
3.4.4 Scoring .....	33
3.4.4.1 Analytic rubric.....	33
3.5 Procedure .....	33
3.6 Questionnaires and interviews .....	36
3.7 Analysis of the data .....	37
 Chapter IV. Results .....	 40
4.1 Descriptive statistics.....	40
4.2 Reliability .....	46
4.2.1 Placement test.....	46
4.2.2 Inter-rater reliability for writing prompts .....	47
4.3 Correlations among placement test scores, writing test scores, self-	

assessment scores, and measures of overall English proficiency .....	51
4.4 Analysis of Variance(ANOVA) .....	53
4.4.1 Planning effect on the Composite Writing Test Scores .....	53
4.4.2 Planning effect on the Analytic Scores .....	55
4.5 Analysis of Covariance (ANCOVA) .....	61
4.5.1 Proficiency mediating influence on effects of planning conditions on writing scores .....	61
4.6. Raters' post-rating feedback .....	63
4.7 Questionnaires and interviews .....	64
4.8 Analysis of learners' writing plan behavior .....	69
Chapter V. Discussion.....	71
5.1 Effects of within-task planning conditions on the analytic and composite scores.....	71
5.2 Korean EFL test-takers' proficiency mediating influence on effects of planning conditions on writing.....	74
5.3. Planning patterns and behaviors of Korean EFL test-takers .....	75
Chapter VI. Conclusion .....	78
6.1 Conclusions and implications.....	78
6.2 Limitations and future studies .....	79
References .....	82
Appendices .....	88
국문초록 .....	114

## List of Tables

Table 2.1 Summary of tests .....	15
Table 3.1 Demographic and TEPS score data of the Korean EFL participants....	25
Table 3.2 Data from the final pool of participants .....	26
Table 3.3 Summary of planning conditions.....	31
Table 3.4 Data from the final grouping of participants .....	35
Table 4.1 Descriptive statistics of pretest and posttest .....	41
Table 4.2 Pretest and posttest score differences in terms of the analytic and composite scores of writing tests among groups.....	42
Table 4.3 Descriptive statistics of the results of participants' self-evaluation on writing ability .....	42
Table 4.4 Descriptive statistics of pretest and posttest according to proficiency level .....	45
Table 4.5 Reliability coefficients for placement test .....	47
Table 4.6 Descriptive statistics of writing prompts showing rater results.....	48
Table 4.7 Spearman's rank order correlation coefficients between raters for two writing prompts as indices of inter-rater agreement.....	49
Table 4.8 Score agreement rates and Kappa coefficients between raters for two writing prompts .....	50
Table 4.9 Pearson correlation coefficients between tests .....	52
Table 4.10 ANOVA for composite scores of writing tests.....	54
Table 4.11 ANOVA for development of idea scores of writing tests.....	55
Table 4.12 ANOVA for organization scores of writing tests .....	56
Table 4.13 ANOVA for vocabulary scores of writing tests .....	57
Table 4.14 ANOVA for sentence variety and construction scores of writing tests.....	58
Table 4.15 ANOVA for grammar scores of writing tests.....	59
Table 4.16 ANOVA for mechanics scores of writing tests .....	60
Table 4.17 ANCOVA for proficiency mediating effect .....	62
Table 4.18 Post-questionnaire data.....	68



## **List of Figures**

Figure 2.1 The Hayes-Flower writing model .....	8
Figure 2.2 Interdependence of language, knowledge and memory processed during writing .....	10
Figure 2.3 Task based performance and language testing .....	11
Figure 2.4 Types of task-based planning .....	17
Figure 2.5 NEAT (National English Ability Test) Level 2 writing question sample .....	19
Figure 3.1 General procedure .....	35
Figure 4.1 Composite mean scores across groups .....	54
Figure 4.2 Development of ideas mean scores across groups .....	55
Figure 4.3 Organization mean scores across groups .....	56
Figure 4.4 Vocabulary mean scores across groups .....	57
Figure 4.5 Sentence variety and construction mean scores across groups .....	58
Figure 4.6 Grammar mean scores across groups .....	59
Figure 4.7 Mechanics mean scores across groups .....	60

# **CHAPTER I**

## **INTRODUCTION**

### **1.1. Background and motivation**

A considerable body of research on the effects of pre-task planning on oral performance has shown that pre-task planning influences both the content of learner's speech and the quality of the language they produce (Foster & Skehan, 1996; Skehan & Foster, 1997; Wendel, 1997; Mehnert, 1998). Some studies reported the beneficial effects of planning on speaking performance (Wendel, 1997; Kawauchi, 2004), but others indicated that there can be an essential trade-off between fluency on the one hand and either complexity or accuracy on the other (Foster & Skehan, 1996, Mehnert, 1998).

However, there have been relatively few studies conducted on the effects of planning on writing. Moreover, most previous studies aimed to identify the effects of pre-task planning which is also called strategic planning. One study investigated the effects of within-task planning on L2 learners' narrative productions along with pre-task planning (Ellis and Yuan, 2004a). Within-task planning refers to the planning that takes place on-line. In other words, it includes planning activities that writers engage in while they are composing their response. Within-task planning can be further differentiated according to the extent to which the task performance is pressured or unpressured. Since within-task planning conditions are similar to the current large-scale English proficiency writing test condition where test-takers are asked to finish their writing within certain time constraints, more empirical research is needed to

investigate the effects of planning conditions in the assessment context.

The issue of planning, in particular, has been much discussed among writing researchers advocating cognitive modeling of the writing process. Writing is described as a dynamic process in which writers must exercise a number of skills and meet great demands at once in terms of task environment. Planning is believed to have three substages: generating; organizing; and goal setting (Hayes & Flower, 1980). This means writers build a representation of knowledge that will be used in writing during the planning stage.

Most recently, many writing researchers have conceptualized planning as an important task condition in writing. According to a model of task-based performance in relation to language testing, the task that is involved and the condition under which it is completed should be carefully considered (Tavakoli and Skehan, 2005). There has been much discussion among writing experts as to the way tasks are implemented in testing within the testing literature. Conditions of task implementation are likely to be a major influence upon how performance can be predicted. Therefore, planning conditions for using writing assessment tasks as tests are worth considering.

In terms of tasks, and in particular writing tasks that require reflection and planning such as academic essays, a question of interest is the optimum time to provide test-takers so they have enough time to plan, write, and revise their writing (Weigle, 2002). This study assumes that the role of planning can indicate the importance of testing conditions and is also closely associated with time limit. The impact of planning conditions clearly warrants investigation, along with other task qualities which may influence language

production in language testing contexts.

Notwithstanding the importance of planning and task conditions, there are no writing tests which require test-takers to make plans while they are writing and examining their plans. Some efforts have been made to use guided planning that encourages test-takers to fill in the blanks and set out writing under strategy-based instruction. Yet, it is still the product, not the process of writing that is observed. Testers are interested primarily in the product and not in the process of writing. Even previous studies that investigated the effects of planning only looked into the learners' final text of writing.

Meanwhile, there have been efforts to develop a rubric to rate writing plans and examine the relationships between the quality of writing plans and writing scores (Chai, 2006). It was meaningful that writing plans were carefully scrutinized, but the study based on only L1 samples and some idiosyncratic features of writing plans were not considered. This current study mainly targeted Korean EFL (English as a Foreign Language) test takers' writing samples, and their behaviors were carefully explored.

Moreover, the mediating effects of Korean EFL learners' proficiency level on planning were taken into account. A considerable amount of research suggests that low-proficiency writers tend to be context-free, intuitive, and plan less and review more at the sentence level. High-proficiency writers, however, tend to plan more, revise more at the discourse level and showed more commitment to the given assignment (Raimes, 1985, 1987; Torrance, 1996; Zamel, 1982, 1983). Investigating how learners' levels of proficiency influence their planning behavior, and how much learners perceive and use

planning can provide valuable insights about test-takers' planning performed during writing tests.

Following these studies, the present investigation intends to further explore the following issues related to the effect of within-task planning conditions on Korean EFL test-takers' performance on argumentative writing assessment tasks.

The main purpose of the current study is to examine the effects of two different planning conditions (i.e., pressured and unpressured) on Korean EFL test-takers' performance on argumentative writing tasks. Not only the composite scores but also six analytic scores are examined. Unlike the previous studies, which attempted to analyze written products in terms of fluency, complexity, and accuracy, the current study uses an analytic rubric consisting of a total of six criteria, which deal with major components of specific writing dimension, which is argumentative writing assessment tasks.

In addition, the present study explores whether Korean EFL test-takers' proficiency mediates the effects of planning conditions on writing scores. This study also investigates how the Korean EFL test-takers perceive and use planning for their argumentative writing assessment tasks. Their planning patterns and behaviors were closely investigated by analyzing their background information and data from questionnaires.

## **1.2 Research Questions**

In this study, Korean EFL learners' writing performances were examined through the composite of the six analytic scores (i.e. development of ideas, organization, vocabulary, sentence variety and construction, grammar and usage, and mechanics). This study intends to answer the following research questions:

1. Do planning conditions have a significant effect on the Korean EFL test-takers' performance on argumentative writing at each of the analytic score levels?
2. Do planning conditions have a significant effect on the Korean EFL test-takers' performance on argumentative writing at the composite score level?
3. Does Korean EFL test-takers' proficiency level mediate the impact of planning conditions on the writing scores?
4. Are there some significant patterns or behaviors of Korean EFL learners' use of planning on argumentative writing depending on their proficiency level?

## **1.2. Organization of the Thesis**

The current study is organized into six chapters. Chapter 2 summarizes the theoretical background and empirical findings about the effects of planning. In Chapter 3, the method of the study is described, which includes participants, raters, examiners, instruments, and an analytic rubric. Chapter 4 reports the results of statistical analyses conducted on writing scores as well as those of

qualitative analyses on questionnaires. Chapter 5 discusses the major findings of the study in relation to the research questions of this study. Finally, Chapter 6 discusses the findings of the study and the implications of the study along with the avenues for future research.

## **CHAPTER II**

### **LITERATURE REVIEW**

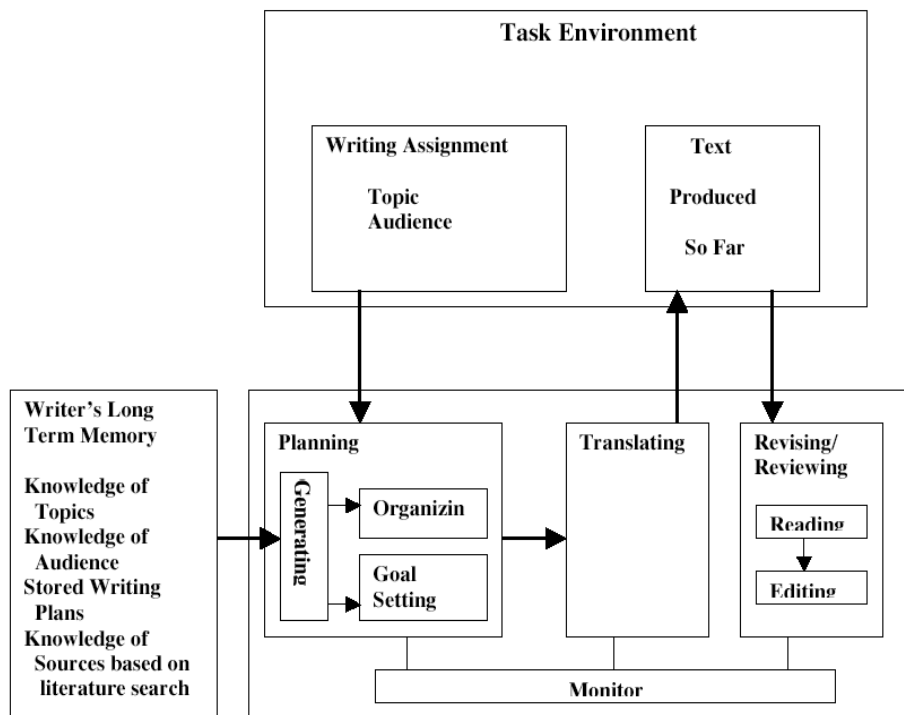
This chapter presents a review of previous studies regarding L2 writing and effects of planning. It begins with a description of a theoretical model of L2 writing and the frameworks of L2 writing assessment design and development, including task characteristics and conditions. The following section discusses the major types of planning used by EFL test-takers and some empirical findings about the effects of planning on their performance on various oral and written tasks. Lastly, this chapter ends by providing research on planning patterns and behavior.

#### **2.1. A Theoretical Model of L2 Writing**

Hayes and Flower's (1980) model has been frequently mentioned in the writing literature, as it provides a common language for discussing cognitive processes in writing. As illustrated in Figure 2.1, Hayes and Flower described the writing process in terms of the task environment, writer's long-term memory and a number of cognitive processes. In the task environment, the writing assignment and the text produced so far are included. The writer's long-term memory works along with knowledge of topic, knowledge of audience, and stored writing plans. Most importantly, a number of cognitive processes, including planning, translating thought into text, and revising are activated. Viewed this way, the writer must exercise a number of skills and meet a number of demands at once.



Writing is a dynamic process in which writers have to deal with an excessive number of simultaneous demands and constraints. It is clear that the writers plan, write, and revise repeatedly, in a way which cannot be divided into clear-cut stages. In the planning stage, writers build a representation of knowledge that will be used in writing. The planning stage has three substages: generating; organizing; and goal settings. Writers generate ideas from memory. Sometimes ideas are well developed and organized in such a way that the writer can formulate them directly in English.



**Figure 2.1 The Hayes-Flower writing model**

However, at other times, the ideas are fragments with little organization. In those instances, writers organize or group ideas and search for subordinate

ideas to develop overall structure. Also, the writers find an order for representing the written text. When it comes to goal setting, the goals are developed by the writer. Some goals are taken directly from memory while others can be developed during writing. During translation, writers work on putting ideas into words under all of the constraints imposed by language, pragmatic, semantic, syntactic, and lexical constraints. The issue of planning, in particular, has been much discussed among writing experts advocating this cognitive modeling aspect of the writing process.

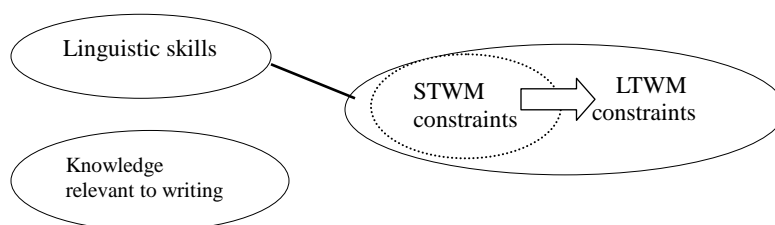
## **2.2 The Frameworks of L2 Writing Assessment Design and Development**

Following earlier L2 writing process studies on rhetorical concerns and composing, Jacobs (1982) made the point that factors beyond linguistic competence determined the quality of students' writing confirming the notion that linguistic competence does not only affect composing competence among second language writers. Zamel (1982) also found that competence in the composing process was more important than linguistic competence in the ability to write proficiently in English. Also, she maintained that when students understood and experienced composing as a process, their writing products would improve. In this study, it is assumed that learners can be encouraged to notice composing as a process through planning and developing their writing products.

Bachman (1990) claimed that in assessments such as writing an essay, the involvement of both language knowledge and the metacognitive strategies is obvious; these include the test-taker's knowledge of grammar, vocabulary,

and organization along with the test-taker's ability to set goals for the essay, to appraise the demands of the task and to plan how to structure the essay. Planning is also viewed as strategic competence since strategic competence is considered to be a set of metacognitive strategies. More specifically, planning refers to deciding how to use what one has. In other words, it is characterized as strategies: selecting elements from the areas of topical and language knowledge for successfully completing the assessment task; formulating one or more plans for implementing these elements in a response to the assessment task; and selecting one plan for initial implementation as a response to the assessment task.

Recently, McCutchen (2011) reviewed linguistic processes that support text production, especially as they interact with other aspects of knowledge relevant to writing, all within the constraints of working memory (Figure 2.2).



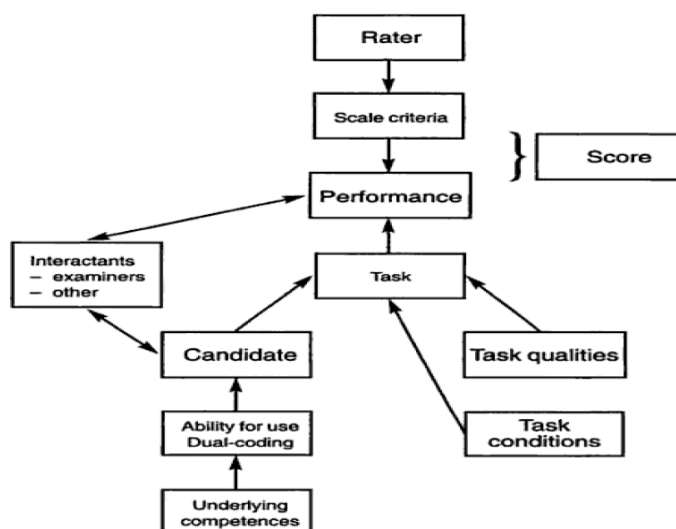
**Figure 2.2 Interdependence of language, knowledge, and memory processes during writing**

Initially, operations of linguistic processes and other processes involving writing-relevant knowledge (e.g., knowledge of genre) are constrained by traditional working memory (or “short-term working memory,” STWM). However, McCutchen suggested that as linguistic skill and writing-relevant knowledge increases, eventually the constraints of STWM give way to more

expansive long-term working memory resources. With regard to this, planning can help ease the load of working memory and enhance the utilization of both linguistic skills and knowledge relevant to writing.

### 2.2.1 General Framework

Tavakoli and Skehan (2005) proposed a model of task-based performance in relation to language testing (Figure 2.3). The model makes it clear that the rating assigned someone on the basis of their performance on a task is the consequence of a whole range of factors. For this reason the task that is involved and the condition under which it is done should be given significant attention. It is important to explore interactions between the components.



**Figure 2.3 Task based performance and language testing**

The testing literature explores various ways in which tasks may be implemented in testing. Conditions of task implementation are likely to be a major influence upon how performance can be predicted. Therefore,

conditions for using tasks as tests are worth exploring. This study assumes that the role of planning can demonstrate the importance of testing conditions.

The impact of planning conditions clearly warrants investigation, along with other task qualities which may influence language production in language testing contexts.

### **2.2.2 Time Limit**

One of the task qualities closely associated with planning conditions is time limit. Weigle (2002) claimed that in writing tasks that require reflection and planning including academic essays, a question of interest is the optimum amount of time to provide test-takers so they are sufficiently able to plan, write, and revise their writing. The question of whether time allocations are appropriate and allow test-takers to perform at their best has been scrutinized; this also refers to how to set the most proper task condition. It is crucial to ensure an appropriate time allowance to maximize test-takers' writing performance through writing assessment tasks.

In the same manner, Bachman (1990) proposed a set of characteristics for describing five aspects of tasks: setting, assessment rubric; input; expected response; and the relationship between input and expected response. One of the characteristics of the expected response is degree of speededness. The degree of speededness is the amount of time that the language user or test taker has to plan and execute a response. In this regard, it is important to give an appropriate time allotment to allow a test-taker to show their writing ability fully.

Although several studies have identified the importance of planning and time allotment, few standardized English tests of writing proficiency allow test takers to have substantial amount of planning time before performing the tasks. The three tests that are most frequently mentioned are compared in Table 2.1.

Some oral performance tests give preparation time and response time separately. Both tasks from TOEFL iBT Speaking test are designed to provide 15 and 30 seconds of preparation time and 45 and 60 seconds of response time. Likewise, two tasks in International English Language Testing System (henceforth, IELTS) consist of 3 to 4 minutes including 1 minute of preparation time each.

On the other hand, most writing tests give time limits combining planning time. Hayes and Grabdwohl Nash (1996) noted that planning and actions are often interwoven in writing. Following the same reasoning, there is no time set aside for planning in current large-scale tests. TOEFL iBT writing test is composed of two tasks and it is stated in the prompt that test-takers have a certain time limit in which to plan, write, and revise their essays. Providing clear instructions for test-takers is obviously an important aspect for a valid test. The amount of detail provided in instructions for writing tests can have an effect on test scores. In this regard, TOEFL iBT writing test seems to consider planning as construct relevant to writing ability. The prompt stated planning explicitly. For a writing test, instructions should be given to maximize construct relevant factors. Similarly, IELTS academic writing test is 60 minutes long with two tasks; one 20 minutes long and the other 40 minutes

long. However, the prompt does not mention planning at all. Cambridge FCE writing test has a relatively long time allotment compared to the other two tests. The prompt does not include any statement about planning but it allows test-takers to notice an appropriate style in writing assessment tasks. Also, the writing test attempts to assess test-takers' abilities to write in different genres or for various purposes. All writing tests have some indication of how long the response should be, requiring a minimum number of words between 150 and 300 words in finished output.

As shown Table 2.1, current writing tests include at least one essay-type task. Argument is definitely the most prevalent genre assigned as a writing task. However, previous studies on planning effects usually dealt with narrative tasks that asked learners to write a composition based on pictures provided (Ellis, 2004a; Ellis and Yuan, 2004b).

Composing argumentative texts presents specific difficulties for writers with respect to content, structure, textual organization, and linguistic coding (Marchand, Coirier, & Dellerman, 1996). Furthermore, the aim of argumentation is not merely to act upon representations of objective knowledge, but mainly on judgments, opinions, beliefs, desires, and subjective preferences. Therefore, learners should understand the writing prompts well and establish a position on the topic in a concise manner. For an argument essay to be effective, it must contain certain elements, that is, an introduction, support, a refutation and a conclusion. Once learners select a position they feel strongly about, they will need to consider logical arguments for the side chosen and the other side to defend against.

**Table 2.1 Summary of Tests**

Test	Type	Task type	Prompt statement	Time limit	Planning time separately	Minimum number of words
<b>TOEFL iBT</b>	Speaking	Independent		45 sec	15sec	none
		Integrative	Read/Listen/Speak	60 sec	30 sec	none
			Listen/Speak	60 sec	20 sec	none
	Writing	Integrative;	You will have... <b>to plan</b> and write your response	20 mins	none	150
		Independent;	You have... <b>to plan</b> , write, and revise your essay	30 mins		300
<b>IELTS</b>	Speaking	Introduction and interview; individual long turn; and two-way discussion		2~3 mins	1 min	none
	Writing	Description of visual information	Summarize the information...and make comparisons where relevant	20 mins	none	150
		Argument	Give reasons...and include any relevant examples...	40 mins		250
<b>Cambridge FCE</b>	Speaking	Interview; long turn; collaborative task; and discussion		14 mins	none	none
	Writing	Article; email; essay; letter; report; review; and story	Write your email. You must use grammatically correct sentences.. in a style appropriate for the situation	80 mins	none	120
			Write your answer...in an appropriate style...			120

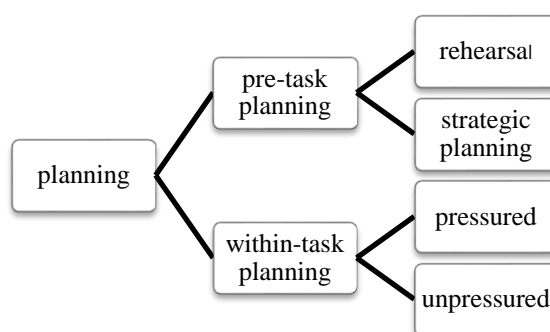


Moreover, finding the appropriate words and making the right connections between words and clauses can be crucial for argumentative writing. Learners should also write sentences which are satisfactory for their level in terms of grammar, syntax, and vocabulary. This means learners have good language skills as well. In order to produce a good final composition, learners should have both writing skills and language skills. Through argumentative writing tasks, learners can show which skills they have or lack.

In addition, it is assumed that writing plans can show the test-taker's visual preparation before they jump into the writing. If learners make a good outline with organized structure, retrieving appropriate words they use in the planning stage, it is expected that learners can produce an effective final text. This study probes into the effects of planning on argumentative writing.

### **2.2.3 Types of Planning**

When it comes to planning conditions, Ellis (2004b) distinguishes two principal types of task-based planning- *pre-task planning* and *within -task planning*, as illustrated in Figure 2.4. These are separated simply in terms of when the planning takes place- either before the task is performed or during its performance. The current study is focused on the latter, which is within-task planning, also referred to as on-line planning. Within-task planning has to do with the planning activities that the writers engage in while they are composing their writing. Within-task planning can be further distinguished according to the extent to which the task performance is pressured or unpressured.



**Figure 2.4 Types of task-based planning**

The pressured condition can be achieved by manipulating the time made available to the learners for on-line planning of what to write in a task performance. However, this condition can be the most demanding as learners have limited processing capacity and are likely to experience difficulty in accessing and encoding their linguistic knowledge. The reverse is the unpressured condition, when learners are required to start performing the task and are given as much time as they wish to take for planning. Researchers believe that unpressured within-task planning may prove beneficial to formulation and also afford time for controlled processing required for monitoring. In this sense, test-takers can freely use their time. Both of these conditions are believed to ease the processing burden (Ellis, 2004b). The current study set out to examine the effects of these two conditions.

#### **2.2.4 Planning Patterns and Behaviors**

To ensure the usefulness of planning, learners' use of planning time and the planning behaviors and strategies were closely examined. Glynn, Britton, Muth, and Dogan (1982) monitored two basic methods of planning called

drafting: one where the writer plans his or her text by creating an outline before writing; the other where the writer begins to write, developing his or her text through discoveries made during the writing process. In their study, they tracked four drafting strategies: organized sentences, which are similar to rough drafting; unorganized sentences like multiple drafting; organized notes, equated to outlining; and unorganized notes. While their results confirm that developing an outline before writing yields the most successful texts, they suggested that individual differences might influence successful use of these four strategies.

In this study, participants' planning behavior, particularly in relation to how they use planning time and how they apply writing strategies can be revealed through questionnaires and interviews. Moreover, Korean EFL learners' sheets used for planning are analyzed.

L2 writers' process strategies, especially the differences between those of more- and less-skilled writers, have been extensively studied (Cumming, 1998; Grabe & Kaplan, 1996). Researchers have commonly found that (a) skilled L2 writers tended to plan more, revise more at the discourse level, and spend more time exploring the most appropriate ways to solve a given task (e.g., Cumming, 1989; Raimes, 1987; Roca de Larios, Mari'n, & Murphy, 2001; Roca de Larios et al., 1999; Zamel, 1982, 1983); (b) unskilled L2 writers tended to plan less and revise more at the word and phrase level (e.g., Raimes, 1985, 1987; Roca de Larios et al., 2001; Roca de Larios et al., 1999; Zamel, 1983); and (c) writers' L1 use, attention patterns, and problem-solving behaviors while writing differed according to such variables as their L2

proficiency, their L1 writing expertise, and the type of tasks they were involved in (e.g., Cumming, 1989; Wang & Wen, 2002).

Since strategic-based instruction and planning have drawn more attention in writing, the recent NEAT (National English Aptitude Test) includes a sample question based on planning in a level 2 actual test. As observed from Figure 2.5, it requires test takers to fill in the blank in a planning sheet and construct a composition based on the planning they made.

The following list shows advantages and disadvantages of a paperbook. Pick one side and write an introduction, then add one more reason to support your position composing the body with three reasons. Lastly, reach a conclusion to finish the argument.	
Advantages Introduction ↓	Disadvantage Introduction ↓
1. not requiring any electronic devices 2. easier to make notes 3. _____	1. heavy to carry 2. taking up too much space 3. _____
↓ Conclusion	↓ Conclusion

**Figure 2.5 NEAT (National English Ability Test) level 2 writing question sample**

The underlying assumption is that planning definitely helps and guides learners' writing. In addition, some formats which consist of three parts, namely, introduction, body and conclusion with some cohesive devices or linking words (e.g. first, moreover, therefore) have already been acquired and frequently used as an exercise or practice in writing instruction. With regard to this strategy and practice for writing, the current study aims to see how Korean EFL learners make use of planning for their writing.

## **2.3. Previous Studies on the Effects of Planning in an Assessment Context**

### **2.3.1 Speaking**

There are several oral studies which provided mixed results about the effects of planning. Several planning studies reported trade-off effects between complexity and accuracy (Foster & Skehan, 1996; Mehnert, 1998). Due to the constraints in capacity during tasks, it is hypothesized that different aspects of language compete for limited attention during processing (Ellis, 2004b; Skehan 1998). As a result, one aspect of performance is prioritized and improves, whereas another aspect receives less attention and remains the same or gets worse. Such phenomena are called trade-off effects (Skehan & Foster, 1997; Yuan & Ellis, 2004).

Wendel (1997) attempted to look into the effects of an opportunity to plan a discourse prior to speech production. The results displayed that fluency and complexity of spoken narrative are greater in the planned condition.

More recently, Kawauchi (2004) investigated the effect of strategic planning on L2 learners' performance on an oral narrative task. Japanese learners of English with different proficiency levels participated and the study also examined the effect of learners' proficiency. The results showed that strategic planning had beneficial effects on fluency, complexity, and accuracy on Japanese learners' oral narratives. The study also concluded that the High EFL group benefited most from the opportunity to plan in the case of fluency and complexity, while the Low EFL group did so in accuracy.

In addition, there have been some efforts to investigate planning effects on speaking in a testing context. Wigglesworth (1997) researched the effects of

planning time in the context of language testing. Using just one-minute pre-task planning, it was shown that performance was improved but this effect was mediated by the difficulty of the task undertaken and the proficiency level of the test-takers. Elder and Iwashita (2004) also examined planning effects of oral performance under testing conditions. They tried to find strategic planning time effects on the features of the oral discourse by test-takers. 197 students consisting of 75 males and 122 females took two tests: a multi-task test of their proficiency, followed by an Institutional version of the TOEFL. Performance of all subjects were rated using analytical rating scales for fluency, accuracy and complexity. The results showed that there was no effect for planning time on performance. These previous studies examined pre-task planning effects on oral performance. However, the effects of within-task planning on written L2 production as well as oral production have not been systematically investigated.

### **2.3.2 Writing**

There are just a few studies examining the effects of within-task planning on writing output. Unlike oral research, planning effects on written performance in a testing context were not investigated. One major study was done by Ellis and Yuan (2004b). They found some beneficial impacts of pre-task planning and within-task planning on the fluency and grammatical complexity of narratives composed by 42 Chinese students of English as a foreign language. The study examined the effects of three kinds of planning conditions (pre-task planning, unpressured on-line planning, and no planning) on written as well as

oral performance. Participants were asked to perform narrative writing tasks which required them to write a story based on two different sets of six pictures. Then, the textual products of the two tasks from three conditions were analyzed in terms of fluency, complexity and accuracy. The findings suggest that the two types of planning had effects on different aspects of L2 writing processes, with pre-task planning promoting formulation and unpressured on-line planning providing better opportunities for monitoring. Writers in the no-planning condition were faced with the need to formulate, execute, and monitor under pressure, with negative consequences for the fluency, complexity, and accuracy of the written product in comparison to that of the planning groups.

However, the length of time spent in completing the tasks took longer in the unpressured on-line planning group and thus, the total number of syllables produced was definitely greater in that condition. Since there was no time constraint on the tasks, it was not clear that longer texts resulted from the effect of careful planning or the length of time spent itself. Ellis and Yuan's (2004) study implies important aspects of the effects of planning. The key to understanding the results lies on learners' limited procedural ability in the L2. They experienced problems in formulating messages and tried to compensate for this lack of procedural ability by monitoring their output using L2 knowledge only if their working memories were not overloaded. Therefore, the opportunities to plan and write helped them to conceptualize, formulate, and monitor by means of controlled processes. In particular, those L2 learners could access their linguistic resources through planning.

Nevertheless, learners' written productions were evaluated with quantifiable measures of accuracy, fluency, and complexity in the study. The total number of words participants formulated, the total number of different grammatical verb forms used in the task along with T-units, and the proportion of clauses that did not contain any error were calculated to analyze the writing. Therefore, the analysis of the written data remained at the sentence-level rather than at the discourse-level, which reflects only language development not considering any coherence of the text itself. To overcome this, the current study attempts to assess the participants' written products by utilizing an analytic rubric which reflects vital subcomponents of a composition. An analytic rubric is expected to give more information on the quality of Korean EFL test-takers' writings in terms of language development as well as writing skill improvement.

Overall, these previous research studies have provided significant insights into the effects of planning on writing. Some studies focused on the effects of planning on narrative writing under pre-task planning conditions, while others concentrated on learners' writing strategies. There are still, however, many more questions that remain unanswered, particularly regarding argumentative writing tasks under within-task planning conditions in a testing context. Since there have been attempts to incorporate planning into test conditions, it is high time that the effects of within-task planning conditions on argumentative writing are finely examined



## **CHAPTER III**

### **METHOD**

This chapter deals with the methodology used to collect and analyze the data. It will begin with the description of participants, raters, and examiners, followed by the instruments used for the placement test and main study: the pretest and posttest. This chapter also provides information on how participants' writing samples were rated. Finally, data collecting procedures and data analyses are described in detail.

#### **3.1 Participants**

Participants for the study were undergraduate students at a major university in Seoul, all of whom were recruited from an online posting according to their self-reported scores of the Test of English Proficiency developed by Seoul National University (henceforth, TEPS). All participants were native speakers of Korean. According to the score band descriptors from Language Education Institute, TEPS scores of 801 or above are equivalent to a level of 1 or 1+ and scores of 600 to 800, to a level of 2 or 2+. These scores are also equivalent to the Test of English as a Foreign Language (henceforth, TOEFL) scores of 107 or above and 86 to 106, respectively (TEPS, 2009). Participants within the former score range in this study are classified as advanced while participants within the latter range are the intermediate group. TEPS scores served as an initial criterion to separate participants into two different levels of groups.

Thirty-one Korean EFL learners participated in the initial part of the present

study. Among the 31 participants, 13 belonged to the advanced group and 18 were subcategorized into the intermediate group depending on their TEPS scores. However, three participants dropped out of the study and were therefore excluded from final calculations, since they did not complete all the tasks needed for the research. Thus, the initial subject pool (n=31) was reduced to 28 participants. Their majors varied from business, education, law, and humanities, to engineering and pharmacy. At the time of the data collection, their ages ranged from 21 to 30.

All personal information was collected through a questionnaire (see Appendix 1). Forty-two percent of the participants responded that vocabulary was their problem area in writing while 35% picked grammar as their weakness in writing. Also, the questionnaire results showed that 17% of the participants had trouble with organization and another 17% had difficulty in content when it comes to writing. Table 3.1 shows the demographic data of the participants for the experiment.

**Table 3.1 Demographic and TEPS score data of the Korean EFL participants**

Subjects		TEPS Score	
Proficiency	N	Mean	SD
Advanced	13	871	54.32
Intermediate	15	705	52.23
Total	28	782	99.34

Since all the undergraduates at the university are required to take TEPS, all students can report scores from the same large-scale standardized English proficiency test. Moreover, TEPS assesses language skills through four

components- listening, grammar, vocabulary, and reading- which gives an estimate of each participant's communicative language skills and English proficiency. Nevertheless, this test lacks a writing component.

Therefore, all participants took a writing placement test to confirm their English writing ability. An adapted version of the Diagnostic Test for Writers by Pearson Education was used for this purpose. The test consists of three parts. Part 1 deals with spelling, capitalization, and punctuation. Part 2 includes items about verbs, pronouns, adjectives, and adverbs. Part 3 covers sentence structure, sentence punctuation, and clarity.

There are a total of 100 items, each with four possible answers and a 90-minute time limit. It is a computer-based test module. However, the placement test for the current study was paper-based. Twenty questions were randomly chosen from the test. Since this test mostly focused on error recognition and error correction with multiple choice questions, 5 short answer questions were added to the test to assess actual writing skills. In this regard, the placement test was designed to assess productive skills of writing as well as receptive skills. There were a total of 25 questions in the placement test and participants were given 15 minutes to complete the test.

**Table 3.2 Data from the final pool of participants**

Participants		Placement test score		TEPS score	
Proficiency	N	Mean	SD	Mean	SD
Advanced	13	26.73	1.37	851.85	85.57
Intermediate	15	20.50	2.73	722.20	66.29
Total	28	23.39	3.84	782.39	99.34

The results of the placement test and TEPS scores were used together to classify the participants into either the advanced or intermediate level. This was done to improve the classification. Table 3.2 shows the data from the final pool of participants.

### **3.2 Raters**

Three raters participated in this study. One was a native English rater from North America and the others were native Koreans. The three raters scored all of the writing samples which were obtained from the pretest and posttest. All raters had extensive experience and knowledge in the fields of ESL/EFL assessment. One worked as an item writer and rater for the TEPS testing program for 4 years. Another is an incumbent English teacher with a near-native level of English proficiency who graduated from a graduate school of TESOL. The third is the researcher of the current study, who was brought in for cases where raters disagreed in their scoring of writing samples. All have rated various writing samples of Korean EFL learners before.

The raters were trained by the researcher on the CBT TOEFL writing analytic rubric. The CBT TOEFL rubric from ETS Research Report by Lee, Gentile and Kantor (2008) was applied to the current study. The rubric was first given to the raters for them to study, review, and become familiar with the scale. Then, the raters received several practice writing samples in testing situations taken from the CBT TOEFL writing. The researcher discussed and tried to resolve some difficulties and discrepancies which raters faced. Raters' scoring for each subcategory was discussed until they reached a consensus

about how to interpret the descriptions of scales and levels. Finally, the two closest pair of scores were selected and used for analyses.

### **3.3 Examiners**

Two Korean examiners took part in this study to guide participants through the tasks. One was the researcher and the other was a graduate student from the same department. Examiners made sure that all laptop computers worked well and participants understood the instructions, giving out a sheet of paper to use for planning and then collecting the planning paper. They also checked and noted the time spent by participants in unpressured within-task planning groups using the *ttclock* timer program. The *ttclock* timer program functions as both a stopwatch and timer, so it can be used to set the time limit on writing and count the time spent on planning. After the tests, the examiners wrote down the time recorded by the *ttclock* program on the back page of the participants' planning sheets.

### **3.4. Instrument**

This section deals with all the instruments used for data collection. First of all, a placement test administered to assess participants' actual writing skills was introduced. Writing tests adapted from the computer-based TOEFL test were used, so the writing prompts and basic instructions were kept the same. However, after taking the same pretest, three groups of participants took the main test under different conditions to examine the effects of within-task planning. The writing samples for each participant were scored using an

analytic rubric.

### **3.4.1 Placement test**

A simplified version of the Diagnostic Test for Writers by Pearson Education made up most of the test. The diagnostic test consists of three parts. There are 11 questions in Section 1 and 9 questions in Section 2. Section 1 covers homonyms, spelling, apostrophes, subject-verb agreement, pronouns, adjectives, and adverbs. Section 2 deals with spliced and fused sentences, fragments, pronoun references, and dangling and misplaced modifiers. All of those questions were chosen from the original version of the Diagnostic Test for Writers by Pearson Education which was provided online. Section 3 has 5 short-answer writing questions which require test takers to infer a thesis statement, add supporting details, and describe pictures. Each question from Sections 1 and 2 is worth 1 point while short answer questions are assigned 2 points respectively (see Appendix 3).

In terms of grading, multiple choice questions were scored straightforwardly by an individual and short answer questions are rated by a holistic rubric ranging from 0 to 2, a scoring method that was adapted from Purpura (2004). Short answer questions were rated by two raters. One was a graduate student from the same department and the other was the researcher. All the scores from raters showed perfect agreement.

### **3.4.2 Writing tests**

Participants were asked to write essays on the topics chosen from a sample writing topic list in the computer-based TOEFL test. According to the pilot test conducted before this study, the two topics assigned to participants were equal in difficulty. Questionnaires also indicated that the participants were quite familiar with the topics, so even intermediate level participants could easily access the given topics. Raters from the pilot study agreed with one another on prompt difficulty.

With the increasing use of computers in testing, this study also required participants to use computers while writing, however, they were allowed to use a paper and pencil when they made notes and plans. Also, *Notepad* which is a text editor for Windows was used instead of using a word processing program that could help participants to easily spot their errors. Furthermore, unlike the word processing program, *Notepad* does not have the function of checking the number of words. All the participants were asked to perform the tasks in a computer-based writing environment to emulate the computer based TOEFL writing test conditions. Also, they were assigned the same prompts to produce an argumentative writing sample (see Appendix 4). On the other hand, the order of tasks was counterbalanced across all the participants to control potential prompt difficulty differences between the pretest and posttest.

### **3.4.3 Planning conditions**

In this study, three experimental conditions were created: no planning; pressured within-task planning; and unpressured within-task planning (Table

3.3). The participants performed the task in the same way as in the condition under which they took the computer-based writing test such as TOEFL CBT test. The researcher and the examiner were present.

**Table 3.3 Summary of Planning Conditions**

<b>Planning condition</b>	<b>Length of time</b>	<b>Response time</b>	<b>Planning time</b>
No planning	30 mins	30 mins	No
Pressured within-task planning	30 mins	25 mins	5 mins
Unpressured within-task planning	30 mins	Remaining time after planning within 30 mins	Unlimited on-line planning time within 30 mins

#### **3.4.3.1 No planning**

Participants were required to write as soon as they received the task. They were not permitted to use a pencil and paper for taking notes or making a plan, but were instructed to produce at least 200 words within 30 minutes. Examiners checked their behavior.

#### **3.4.3.2 Pressured within-task planning**

Participants were asked to make plans for 5 minutes and then finish writing within an additional 25 minutes. They were also instructed to produce at least 200 words. In this way, the participants were pressured to perform the task with limited opportunities for on-line planning.

In the case of oral production, several studies (Foster & Skehan, 1996; Wendel, 1997; Mehnert, 1998) showed that when at least a 10-minute planning time was provided, there were measurable effects on all three aspects



of language use—fluency, accuracy, and complexity. However, this turned out to be too long for Korean EFL learners in the case of writing. A pilot study involving 50 advanced Korean EFL learners was carried out to determine the length of planning time for the three experimental conditions in this study. Almost all of the learners took less than 5 minutes for planning. Therefore, 5 minutes was determined to be sufficient for planning and was used for this study.

No detailed guidance was provided, but the participants were asked to plan their writing as they usually do. The participants were provided a sheet of paper to write notes. Participants were allowed to keep the notes as they were writing since they could be used as a reference. The notes also provided as evidence regarding how Korean EFL learners undertook the planning and could be referred to in the interview.

#### **3.4.3.3. Unpressured within-task planning**

Participants were required to make plans before they wrote. As they had full control of managing their time, they could make use of planning time on their own. In other words, they could take as long as they wanted to plan. However, it also meant that the more time they spent on planning, the less time they had in which to write. The participants were instructed to click the *ttclock* timer button on the screen as soon as they finished planning. The *ttclock* timer and stopwatch program measured the time learners spent on planning and the examiners noted the time. Participants were also required to produce at least 200 words, the same as on the pressured within-task planning task. Like in the

actual TOEFL writing session, the time limit and essay length requirements were designed to push participants to write rapidly.

### **3.4. 4 Scoring**

#### **3.4.4.1 Analytic rubric**

An analytic scale was applied to pretest and three conditioned tests. The analytic rubric was adapted from ETS Research Report by Lee, Gentile and Kantor (2008) for assessing TOEFL CBT essays (see Appendix 6). In analytic (or multitrait) scoring, writing samples are rated on several important aspects of writing quality, rather than being assigned a single overall rating (Weigle, 2002). An analytic scale is more detailed than a holistic rubric, consisting of either five or six bands for six criteria: development of ideas, organization, vocabulary, sentence variety and construction, grammar and usage and mechanics. It covered all possible components of writing which can provide more useful information about participants' writing skills and language skills. The final composite scores combined all six criteria, and thus the maximum score possible is 32.

There are a number of advantages to analytic scoring. Most importantly, it resolves the problem of uneven development of sub-skills of writing in individuals. Also, the fact that the scorer has to give a number of scores will tend to make the scoring more reliable (Weigle, 2002).

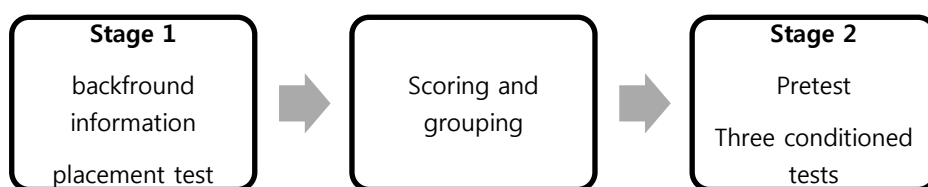
### **3.5 Procedure**

As illustrated in Figure 3.1, the data collection was carried out in two stages.

During Stage 1, the participants filled in a questionnaire regarding their background information. It was mainly about participants' demographic information, such as their age, academic field or major, length of residence in English-speaking countries and other useful information (see Appendix 1). Moreover, a questionnaire about their writing strategies was administered. The questionnaire consists of four parts: general learning strategies; general writing strategies; before writing strategies and during writing strategies. Each part comprises 7 questions using a 5- point Likert scale.

Afterwards, the participants took a placement test to confirm their English writing abilities. The test consisted of 25 questions, and the participants had 15 minutes to complete the test. According to the scores from the placement test, they were classified into two groups again: advanced and intermediate. Then, they were randomly assigned into three experimental groups: no planning; pressured within- task planning; and unpressured within-task planning. Table 3.3 shows the data from the final grouping of participants. The PP group had a relatively higher mean value of TEPS scores than those of the Control and UnPP groups. Therefore, one-way ANOVA was performed to check whether the differences in the three groups' TEPS scores were statistically significant. No statistically significant differences were found among the three groups in the TEPS scores ( $F=.912, p>.05$ ).

In Stage 2, each participant was required to take a pretest without planning for 30 minutes. All participants were not permitted to make notes. They were instructed to write as soon as they started the task.



**Figure 3.1 General Procedure**

**Table 3.4 Data from the final grouping of participants**

Participants				Placement test score		TEPS score	
Grouping	N	Adv	Int	Mean	SD	Mean	SD
Control	9	4	5	19.55	3.78	768.89	103.00
PP	10	5	5	20.65	3.25	809.00	92.06
UnPP	9	4	5	18.61	4.29	766.33	108.54
Total		28		19.64	3.74	782.39	99.34

*Note.* Control = No planning condition; PP = Pressured within-task planning condition; UnPP = Unpressured within-task planning condition

Adv= advanced, Int= intermediate

The control group took the second test under the same condition. The pressured within-task planning group had 5 minutes of planning time and another 25 minutes was given to complete their writing. The unpressured within-task group took the second test for 30 minutes, but participants in this group were allowed to take as much time as they wanted for planning since the 30 minutes allotted for the task was under their control. Except for the control group, the two groups were provided a sheet of paper and a pencil to work on planning before they used the computer to write a composition. Participants in the two planning groups were told that their planning papers should be submitted with their final compositions. Prompts A and B involved counterbalancing, which presents the conditions in all possible orders to avoid

order effects (Vogt, 2005) and control for the difficulty difference between prompts.

On completion of the tasks, all participants were asked to fill out a post-test questionnaire which was done to examine what they really did while taking the writing test and planning. Lastly, some participants from each conditioned group had an oral interview with the researcher. The interview was conducted in such a way that the researcher followed up on the participants' responses to some of the questionnaire items. Random numbers were given to all participants during the study for scoring purposes. Similarly, for the writing files obtained for tasks, other random numbers were assigned.

### **3.6 Questionnaires and interviews**

All participants were asked to fill out a questionnaire before the placement test and shortly after finishing the two writing tests. The questionnaire consists of two parts: one is self-assessment of English writing strategies (see Appendix 2) and the other is survey questions related to participants' test taking experiences such as appropriateness of test instructions, testing and planning time, the difficulty of writing topics, and other comments about the tests (see Appendix 5).

A simplified version of the English as a Second Language Program 82 Questionnaire; Self-Assessment of English Writing Skills and Use of Writing Strategies by Marquette University was used. The questionnaire consists of four parts; general learning strategies, general writing strategies, before writing and during writing strategies. There are 7 questions in each part and

all of those questions were chosen from the original version of the ESLP 82 Questionnaire by Marquette University which was provided online.

Both types of questionnaire use a 5-point Likert scale, and the latter type questionnaire includes more open-ended questions, for instance, how the participants perceive the planning, how they make use of planning time and whether they had instruction about how to plan before writing. The participants' responses to the questionnaire were used to help interpret the findings of the statistical analysis. This information could provide useful insights into participants' planning behavior.

Moreover, some participants had an oral interview with the researcher immediately after the post-test questionnaire, based on their responses. In this respect, the responses to the questionnaire were finely scrutinized during the interview and verbal feedback on participants' attitudes and behaviors towards planning was also gathered.

### **3.7 Data Analysis**

All scores submitted by the raters were entered into the Microsoft Excel 2007 spreadsheet and sorted out. This was used to process writing scores, compute basic statistics, and create charts for the raw scores of writing. Then the data were transferred to IBM SPSS 20 (Statistical Package for Social Studies) for Windows (IBM SPSS 20, 2011). This software was used to obtain the descriptive statistics for writing scores and correlations among these scores, and other criterion measures and reliability coefficients for rating.

First, one-way ANOVA was carried out to find any significance in pretest

among groups. Then, repeated measures ANOVA for composite scores from the pretest and posttest was conducted since this study examined the effect of planning conditions. Statistical tests with significant results at the alpha level of 0.05 are discussed here. In addition, the practical significance of some findings, as indicated by the effect size measure of Cohen's *d* (Cohen, 1988), is considered to see if the statistically significant difference is large enough to be of value or concern in a practical sense. Moreover, a set of repeated measures ANCOVA for both analytic and composite scores from the pretest and posttest according to learners' proficiency was conducted to see whether there was any mediating effect of proficiency on the impacts of planning conditions on writing scores.

As follow-up analyses, a series of repeated measures of ANOVAs investigating analytic scores were performed followed by post-hoc Bonferroni tests. Since there were three different planning conditions for the independent variable, the means of two and three groups on a dependent variable can be tested simultaneously for significant differences. ANOVA is considered the most appropriate for analyzing the collected data because the experimental design involved participants being assessed across three different planning conditions. In other words, there are more than two mean scores since there are three groups which have three possible comparisons. Also, Bonferroni tests were conducted since they have more power than Tukey when fewer tests are done. It is also good for testing planned comparisons.

In terms of reliability, first of all, Cronbach's alpha, particularly item alpha, is used to compute the score reliability coefficients for Sections 1, 2, and 3 in

the placement test. Pearson correlations, rater agreement indices, and kappa coefficients were calculated to examine the inter-rater reliability. The agreement indices between two raters were also computed. As mentioned earlier, when there was a discrepancy between raters, scores from a third rater were included.

In addition, the relationship among placement test scores, writing test scores, and self assessment scores and measures of English proficiency was examined by correlation. Correlation can be used as an index to evaluate the potential for linking scores from different tests. In this study, it was also used to investigate what relationship existed between scores for each section and the composite writing scores. Pearson correlations were computed among the subsection scores, the total scores, and overall English proficiency measures.

Furthermore, the data from questionnaires and an in-depth interview were closely examined. Scores from the 5-point Likert scale were entered into the Microsoft Excel 2007 spreadsheet which was used to compute the descriptive statistics. An analysis of open-ended questions from the questionnaire and an interview was carried out to supplement the quantitative analysis.



## **CHAPTER IV**

### **RESULTS**

This chapter presents the results of the statistical and psychometric analyses of the data: descriptive statistics; reliability coefficients; inter-rater reliability; and correlation coefficients. For clarity of presentation, this chapter is divided into five sections. First, it examines the descriptive statistics of the writing tests and self-assessment tests. Moreover, it examines the reliability of the placement test and raters. Then, it investigates the effect of planning conditions on the argumentative writing test scores through repeated measures ANOVA and the mediating effect of test-takers' proficiency by ANCOVA. Finally, it presents analyses of the raters' post-rating feedback about the analytic scoring rubric, questionnaires regarding writing strategies, and planning behavior, followed by interviews.

#### **4.1. Descriptive statistics**

Table 4.1 displays the descriptive statistics which include means and standard deviations of both analytic and composite writing scores for three groups. Presenting the analytic scores allows examination of the impact of planning both on writing-and language-related qualities of the participants' essays.

**Table 4.1 Descriptive statistics of pretest and posttest**

Rating dimensions	Pretest*													
	DI		O		V		SV		G		Me		total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	5.00	.500	5.33	.707	4.56	.527	4.44	.882	4.11	.782	4.44	.726	26.56	3.35
PP	4.40	.966	4.50	1.17	4.30	.483	4.20	.789	4.30	.675	4.70	.675	25.50	4.14
UnPP	5.22	.833	4.67	1.00	4.33	.707	4.44	.882	4.00	.707	4.56	.726	26.33	3.57

Rating dimensions	Posttest*													
	DI		O		V		SV		G		Me		total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	4.67	1.00	5.00	.866	4.33	.707	4.44	.726	3.89	.782	4.11	1.05	25.44	3.97
PP	4.70	.823	4.80	.919	4.60	.516	4.50	.527	4.00	.816	4.40	1.07	26.30	3.40
UnPP	4.78	1.30	4.89	1.16	4.22	.667	4.33	.707	4.00	.500	4.56	.726	25.78	3.86

*Note.* \*Prompts A and B were counterbalanced

DI=development of ideas; O=organization; V=vocabulary; SV= sentence variety and construction; G=grammar;  
Me=mechanics

**Table 4.2 Pretest and posttest score differences in terms of the analytic and composite scores of writing tests among groups**

Section	DI		O		V		SV		G		Me		total	
	D	%	D	%	D	%	D	%	D	%	D	%	D	%
Control	-.33	-5.5	-.33	-5.5	-.23	-3.8	0	-	-.22	-3.6	-.33	-5.5	-1.12	-3.5
PP	.30	5	.30	5	.30	5	.30	5	-.30	-5	-.30	-5	.8	2.5
UnPP	-.44	7	.22	3.6	-.11	-1.8	-.11	-1.8	0	-	0	-	-0.55	-1.7

*Note.* D= difference

**Table 4.3 Descriptive statistics of the results of participants' self-evaluation on writing ability**

Section	Grammar		Vocabulary		Organization		Content		Mechanics		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	3.33	1.32	2.78	.66	2.89	.78	3.11	.60	2.56	1.13	11.33	2.50
PP	3.40	.69	2.80	.63	3.20	.63	3.10	.56	3.00	.81	12.10	1.91
UnPP	3.11	.78	2.78	.66	3.22	.66	3.33	.70	3.33	.50	12.67	1.87
tot	3.29	.93	2.79	.63	3.11	.68	3.18	.61	2.96	.88	12.04	2.09

As mentioned in the previous chapter, participants were classified into three different groups: no planning (i.e., control); pressured within-task planning, and unpressured within-task planning group. The mean scores were compared across three different groups. Judging from the means, the three groups showed trivial differences between tests.

The control group had a decrease across all areas and the UnPP group also showed a similar pattern except in the organization section. On the other hand, the PP group showed a modest increase in sections such as the development of ideas, organization, vocabulary, sentence variety and construction. Despite its small value, the PP and the UnPP group showed improvements in the organization section. However, all groups showed a slight decrease or remained unchanged in the grammar and mechanics sections.

Table 4.2 reports the differences in the analytic and composite scores between two tests. Together with the raw scores' differences, it also presents the percentages of change between scores in the pretest and posttest. The composite score of the control and the UnPP group had a decrease of 3.5% and 1.7% respectively, while that of the PP group had an increase of 2.5%. The control group showed declines across all areas apart from the sentence variety and construction area which was not affected. The PP group benefited most among the three groups with an increase of 5.5% in four sections. However, these benefits were cancelled because of a reduction of 5% in two areas. In the case of the organization section, the control group seemed to have a negative effect with a decrease of 5.5%, but the other groups benefitted with an increase of 5.5% and 3.6% respectively.

Accordingly, the descriptive results demonstrated that there were modest differences among the groups in terms of small increases and decreases of composite and analytic scores.

When it comes to self assessment of writing, as it is shown in Table 4.3, the mean of composite scores among all groups was 12.04 which was quite modest considering the highest possible score was 20. The Control and the PP groups scored higher in grammar than the UnPP group, averaging 3.33 and 3.40 respectively. By contrast, the UnPP group rated themselves higher in the content and mechanics dimensions, with 3.3 for each. Nevertheless, all groups graded themselves relatively low scores for the vocabulary dimension.

Table 4.4 displays the descriptive statistics: means of both analytic and composite writing scores for three conditioned groups depending on proficiency; advanced and intermediate. Judging from the means, the three groups showed trivial differences between pretest and posttest.

In terms of advanced learners, the control group had a slight increase in sections such as the development of ideas, organization, vocabulary, and grammar. On the other hand, the PP group showed a modest decrease in sections such as the development of ideas, organization, sentence variety and construction, and grammar. The UnPP group had a slight increase in sections like development of ideas, organization and grammar. In regard to composite scores, the PP group declined while the UnPP had a slight increase despite their small value. Rather, the control group had an increase.

**Table 4.4 Descriptive statistics of pretest and posttest according to proficiency level**

Advanced group															
Rating dimensions	DI		O		V		SV		G		Me		total		
	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	
Control	5	4.75	5.5	5.25	5	4.25	4.75	4.75	4.25	4.25	5	4.25	28.75	27	
PP	4.4	4.6	4.4	4.8	4.2	4.6	4.4	4.6	4.6	4.2	5	5	26	27.4	
UnPP	5.75	5.25	5.25	5.5	5	4.5	5	4.75	4.5	4.25	5	4.75	27.25	28	
Intermediate group															
Rating dimensions	DI		O		V		SV		G		Me		total		
	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	
Control	5	4.6	5.2	4.8	4.2	4.4	4.2	4.2	3.6	3.6	4	4	24.8	24.2	
PP	4.4	4.8	4.6	4.8	4.4	4.6	4	4.4	4	3.8	4.4	3.8	25	25.2	
UnPP	4.8	4.4	4.2	4.4	3.8	4	4	4	3.6	3.8	4.2	4.4	24	24	

Regarding the intermediate learners, the control group had a modest increase only in grammar while the PP group went up in organization and sentence variety and construction sections. The UnPP group had an increase in vocabulary, grammar and mechanics. Nevertheless, the values were so small that there was no significant planning effect on the intermediate learners.

## **4.2. Reliability**

### **4.2.1 Placement test**

Cronbach's alpha, particularly item alpha, is a widely used measure of internal score consistency across items, and was used to compute the score reliability coefficients for Sections 1, 2, and 3 in the placement test. Consistent outcome measures can indicate the validity of the developed test. Generally, alpha equal to or greater than 0.7 is considered "acceptable" for reliability in most research situations (Bland & Altman, 1997). After initial item analysis, one item, which was Question 4, was excluded from Section 1 because of zero variance. Three items, Questions 12, 13, and 16, turned out to have very low item discrimination, so they were eliminated from Section 2 for the calculation. This section provides reliability coefficients for each section and for all sections combined.

While Sections 1 and 2 consist of multiple-choice questions, Section 3 has short answer questions. Therefore, the composite scores of Section 1 and 2 as well as composite scores of Sections 1, 2, and 3 are also provided.

**Table 4.5 Reliability coefficients for placement test**

Sub- division	N	$\partial$	$\partial'$
Section1	10	.66	.65
Section 2	6	.39	.38
Section 1+2	16	.70	.69
Section 3	5	.48	.47
Section 1+2+3	21	.71	.72

*Note.* N=number of test items;  $\partial$ = Cronbach's alpha (reliability coefficient);  $\partial'$ =  $\partial$  based on standardized items

Table 4.5 shows the reliability coefficients for Sections 1, 2 and 3 and for all three sections combined. Internal consistency indices represent whether the test takers responded to the items consistently in a single trial.

The alphas were 0.66 for Section 1, 0.39 for Section 2, and 0.70 for the two sections combined (Sections 1+2). The alpha was 0.48 for Section 3 and 0.71 for all sections combined (Sections 1+2+3). When these were used to compute the reliability for standardized items, their values were changed slightly. These reliabilities represent the consistency of test scores. The composite values of both and all sections were 0. 70 and 0.71, which are acceptable levels of reliability. Cronbach's alpha is expected to increase when more items are added to a test. Since the number of items in the placement test is low because of time constraints, the predicted reliability can be expected to have an upward tendency.

#### **4.2.2 Inter-rater statistics and reliability for writing prompts**

Inter-rater reliability was assessed in several ways. First of all, descriptive statistics of writing tests by the three raters were examined. Then, Pearson correlation coefficients were computed to examine the relationship among



raters. In terms of writing tests, since the scores were based on an analytic rubric, correlations were based on the scores of each rater for each section. Next, the agreement indices between raters were calculated to examine the percentages of perfect, adjacent, and non-adjacent scores between both raters. The final measure of inter-rater reliability was Kappa coefficients.

**Table 4.6 Descriptive statistics of writing prompts showing rater results**

<b>Prompt A</b>												
Sc	DI		O		V		SV		G		Me	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
R1	3.96	1.03	4.00	1.05	3.86	.80	3.79	.91	3.75	.88	3.82	.98
R2	4.75	1.14	4.86	1.00	4.36	.62	4.46	.79	3.93	.66	4.61	.68
R3	4.21	.87	4.39	.73	3.96	.63	4.21	.78	3.89	.73	3.89	.91

<b>Prompt B</b>												
Sc	DI		O		V		SV		G		Me	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
R1	4.68	1.09	4.32	1.02	4.29	.71	4.29	.65	3.86	.84	4.04	.92
R2	4.75	1.04	4.79	1.28	4.36	.67	4.25	.70	3.54	.79	4.43	.95
R3	4.54	.99	4.36	.82	4.25	.58	4.18	.72	3.79	.68	4.25	.96

Table 4.6 shows the means and standard deviations of each section score assigned by three raters. This is the very first step for measuring inter-rater reliability, which is useful in determining whether a particular scale is appropriate or not. In other words, inter-rater reliability minimizes variation and increases validity in the application of a scoring rubric.

As shown in Table 4.5, the mean scores for Rater 2 are higher than those of Raters 1 and 3, but the mean differences either between Rater 2 and 3 or Rater 1 and 3 seem to be relatively small. The values and rating patterns of Rater 1 and Rater 2 were relatively different even though they had careful

training respectively. Moreover, it is likely that their native language and background can influence their rating pattern and severity. Rater 2 is an English native speaker from North America who has a more lenient rating tendency on EFL writing texts while Rater 1 is a native Korean who used to be an English teacher and is more stringent. To be more specific, Spearman's rank order correlation coefficients among three possible pairs of raters were examined.

As observed through the results of Pearson correlations between raters in Table 4.7, there was a higher correlation between Rater 1 and Rater 3. The high correlation indicates that there is stability of scores across the raters. The inter-rater reliability values are very high, near 0.8 and 0.9. This also means that the raters showed a similar rating pattern in terms of the degree of rank ordering essays.

**Table 4.7 Spearman's rank order correlation coefficients between raters for two writing prompts as indices of inter-rater agreement**

<b>Prompt A</b>						
	<b>DI</b>	<b>O</b>	<b>V</b>	<b>SV</b>	<b>G</b>	<b>Me</b>
R1-R2	.31	.42	.35	.62*	.59*	.61*
R2-R3	.70*	.72*	.68*	.81*	.67*	.58*
R3-R1	.77*	.74*	.79*	.76*	.93*	.96*
$\partial$	.83	.83	.80	.91	.79	.72
<b>Prompt B</b>						
	<b>DI</b>	<b>O</b>	<b>V</b>	<b>SV</b>	<b>G</b>	<b>Me</b>
R1-R2	.44	.30	.43	.31	.27	.73*
R2-R3	.60*	.66*	.73*	.63*	.31	.85*
R3-R1	.86*	.76*	.80*	.67*	.92*	.85*
$\partial$	.74	.79	.81	.78	.51	.95

*Note.* \*= significant at 0.01 level (2-tailed);

$\partial$ = adjusted reliability coefficients based on R2-R3 by applying the Spearman-Brown Prophecy Formula

Among the pairs, the median pair of scores, which was Rater 2 and Rater 3, were selected and used for analysis. As there were just two writing prompts to rate, the reliability to compare in relation to the test length, or the number of test items was estimated. The projected test-length was predicted by using the Spearman Prophecy Formula (Spearman, 1910) was used. The predicted reliability gets even higher.

Table 4.8 presents agreement indices as well as Kappa coefficients which serve as further evidence of inter-rater reliability in the scoring of writing tests. The results indicate that the rate of perfect + adjacent agreement ranges from .85 to 1.0 for all subsections of the writing scale. A perfect agreement rate means that the two raters gave the same scores to most of the writings, while the adjacent agreement rate is obtained by computing the percentage of ratings that differ by only one score band between two raters.

**Table 4.8 Score agreement rates and kappa coefficient between raters in two writing prompts**

Prompt	SC	PA		AA		PA+AA		NA		Kappa
		N	Rate	N	Rate	N	Rate	N	Rate	
<b>A</b>	DI	12	.42	13	.46	25	.88	3	.10	.25
	O	12	.42	15	.53	27	.96	1	.03	.22
	V	17	.60	11	.40	28	1.00	-	-	.35
	SV	21	.75	7	.25	28	1.00	-	-	.59
	G	19	.68	9	.32	28	1.00	-	-	.47
<b>B</b>	M	12	.42	13	.46	25	.88	3	.10	.18
	DI	13	.46	12	.42	25	.88	3	.10	.28
	O	10	.35	14	.50	24	.85	4	.14	.21
	V	21	.75	7	.35	28	1.00	-	-	.57
	SV	21	.75	6	.21	27	.96	1	.03	.59
	G	7	.25	21	.75	28	1.00	-	-	-.16
	M	23	.82	5	.18	28	1.00	-	-	.69

*Note.* PA= Perfect score agreement; AA= Adjacent agreement; PA+AA= Perfect score agreement plus adjacent agreement; NA= Non-adjacent agreement

Thus, the scores regarded discrepant are those that differed by 2 bands or more. These score discrepancy rates ranged from .03 to .14, which represent that the two raters had a relatively high agreement rate overall. Vocabulary, sentence variety, and grammar in writing test A had no cases of non-adjacent agreement while vocabulary, grammar and mechanics in writing test B did.

Kappa coefficients<sup>①</sup> were relatively low compared to the high correlations ranging from .18 to .69. However, it is generally thought to be a more robust measure since Kappa takes into account the agreement occurring by chance. In other words, this indicates the level of agreement after chance probability is removed. The values were considerably high in the sentence variety section of both writing tests, 0.59.

#### **4.3. Correlations among placement test scores, writing test scores, self assessment writing scores, and measures of overall English proficiency**

Pearson correlation coefficients were calculated in order to examine the relationship among the placement test, writing tests and measures of English proficiency. The assumption is that high coefficients between the placement test score and TEPS scores and between writing test scores would indicate a close relationship among tests. In other words, this also serves as evidence of the validity of the placement test and writing tests. Moreover, self assessment

---

<sup>①</sup> The Cohen's Kappa Coefficient Index can obtain a grade to compare

K	Interpretation
0.00-0.20	Poor agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Good agreement
0.81-1.00	Excellent agreement

scores were examined. The results of Pearson correlation among the tests are presented in Table 4.9.

**Table 4.9 Pearson correlation coefficients between tests**

TEST		Writing Prompts A+B scores						TEPS scores		
		DI	O	V	SV	G	Me	gr	vo	tot
<b>WP</b>	DI	1								
<b>A+B</b>	O	.81*	1							
<b>scores</b>	V	.58*	.73*	1						
	SV	.67*	.61*	.64*	1					
	G	.47	.59*	.65*	.74*	1				
	Me	.21	.23	.37	.23	.66*	1			
<b>TEPS</b>	gr	.25	.30	.50*	.44	.53*	.35	1		
	vo	.26	.20	.39	.42	.31	.27	.65*	1	
	tot	.24	.69*	.49*	.60*	.41	.49*	.87*	.84*	1
<b>Self</b>	vo	.38	.24	.46	.59*	.41	.46	.65*	.72*	.62*

*Note.* \* = significant at .01 level (2-tailed);

The sum of correlation of writing prompts was conducted by Fisher and Fisher inversion; PT= placement test

When it comes to the placement test, the scores of section 3 which required learners to write short answers were compared to TEPS scores because the questions types of Sections 1 and 2 were similar to those of the TEPS grammar section. The results of Pearson correlation showed that Section 3 had moderate correlations with the TEPS grammar (.62), vocabulary (.54) and total score (.51). The moderate relationship between the placement writing test and TEPS suggests that the placement writing test may not be a strong indicator of general English proficiency as measured by TEPS. Since the placement writing test was designed to measure the writing ability of learners, the moderate relationship means that the scores of TEPS may predict learners' writing ability to some extent.

Among the six rating dimensions of the writing tests, development of ideas has a high correlation with the organization dimension (.81). Grammar also has a high correlation with vocabulary (.65), and sentence variety and construction (.74). The self assessment vocabulary score had a moderate correlation with sentence variety and construction (.59).

#### **4.4. Analysis of Variance (ANOVA)**

##### **4.4.1 Effects of planning conditions on the composite writing test scores**

This section demonstrates the results of repeated measures ANOVAs to observe the interaction between groups and within groups.

First of all, to check whether the differences in the three groups' pretest scores are statistically significant, one-way ANOVA was performed. No statistically significant differences were found among the three groups in the composite scores of the pretest ( $F= 2.15, p>.05$ ). Nor did any dimension show any statistically significant differences.

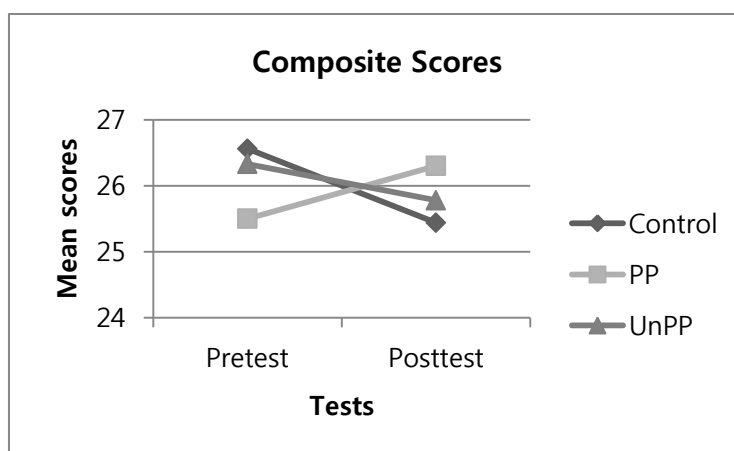
Table 4.10 and Figure 4.1 show the results of repeated measures ANOVA of the composite scores across two different tests. It provides the score differences among the three different groups as well as the differences within the groups under different planning conditions (i.e., no-planning group; pressured within-task planning group; and unpressured within-task planning group).

The results show that there was not a statistically significant difference between groups,  $F=.005, p>.05$ . Also there was no significant Time  $\times$  Type interaction,  $F=1.050, p>.05$ . The mean values of the composite score of the

pretest were 26.56, 25.50 and 26.33 while those of the posttest were 25.44, 26.30 and 25.78. This means that the variance among groups stayed almost the same in both tests despite a small difference of only 1.12, .8 and .55. Since each group was equivalent except for the planning condition, this result implies that there was no statistically significant planning effect on the writing tests.

**Table 4.10 ANOVA for composite scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	.238	.119	.005	.995
Errors	25	586.244	23.450		
<b>Within Groups</b>					
Time	1	1.166	1.166	.264	.612
Time x Type	2	9.269	4.635	1.050	.365
Errors	25	110.356	4.414		



**Figure 4.1 Composite mean scores across groups**

As observed in Figure 4.1, the control group and the UnPP group showed an almost identical downward tendency. The two groups' composite scores declined from pretest to posttest. On the other hand, the PP group displayed an

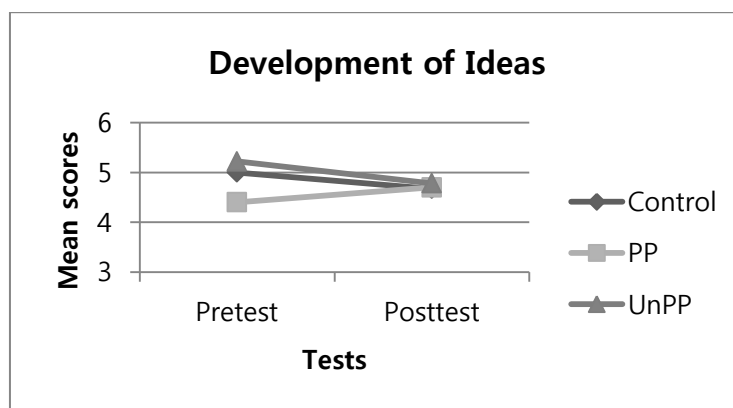
upward tendency.

#### 4.4.2 Effects of Planning Conditions on the Analytic Scores

Even though the results of repeated measures ANOVA in the composite writing scores showed no statistically significant values, it is necessary to further explore the analytic scores from each section respectively. This section reports the results of a series of repeated measures ANOVAs for each of the analytic criteria: development of ideas; organization; vocabulary; sentence variety and construction; grammar; and mechanics.

**Table 4.11 ANOVA for development of idea scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	1.979	.989	.812	.455
Errors	25	30.450	1.218		
<b>Within Groups</b>					
Time	1	.354	.354	.673	.420
Time x Type	2	1.533	.777	1.475	.248
Errors	25	13.161	.526		



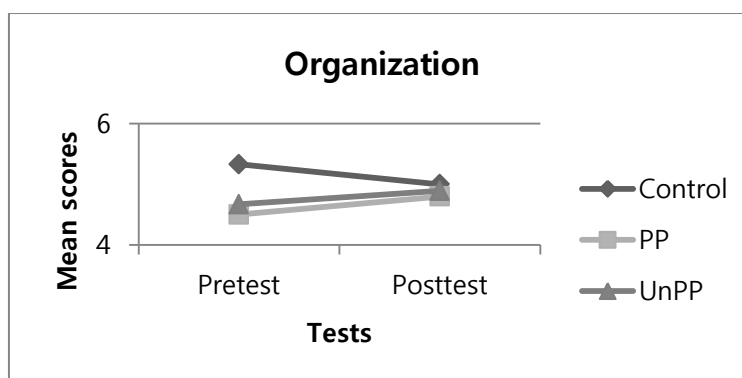
**Figure 4.2 Development of ideas mean scores across groups**



As observed from Table 4.11, similar results were obtained for the development of ideas scores. Even though there was a slight increase and decrease in scores for each group respectively, the results of repeated measures ANOVA indicated that this difference was not statistically significant between groups,  $F=.812$ ,  $p>.05$ . Moreover, there was no statistically significant Time  $\times$  Group interaction,  $F=1.475$ ,  $p>.05$ . However, it was noticeable from Figure 4.2 that the Control and the UnPP group had the same downward tendency while the PP group had a somewhat upward pattern across tests, which was again the same as the composite scores.

**Table 4.12 ANOVA for organization scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	2.696	1.348	.883	.426
Errors	25	38.161	1.526		
<b>Within Groups</b>					
Time	1	.055	.055	.128	.724
Time $\times$ Type	2	1.101	.550	1.271	.298
Errors	25	10.828	.433		



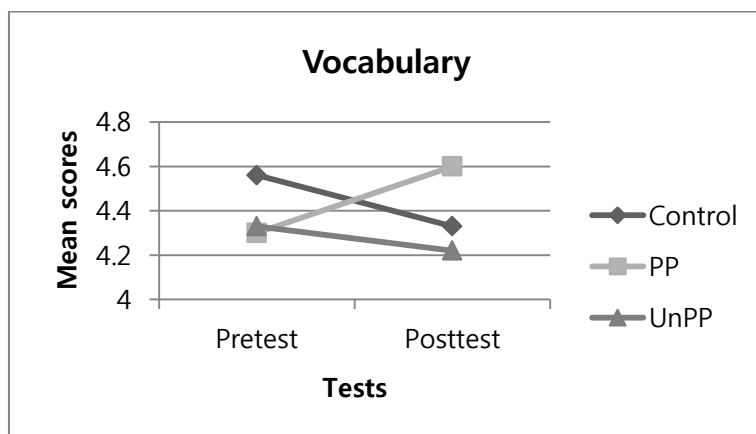
**Figure 4.3 Organization mean scores across groups**

In terms of organization, there were also minimal differences of mean scores between pretest and posttest. The results of repeated measures ANOVA showed no statistically significant values between groups,  $F=.883, p>.05$ .

Also, there was no statistically important Time  $\times$  Type interaction,  $F=1.271, p>.05$ . As observed from Figure 4.3, however, the mean scores from the PP and UnPP group improved slightly. By contrast, the control group had a slight reduction in the posttest score. Unlike the previous section, the PP and the UnPP groups showed a similar upward pattern.

**Table 4.13 ANOVA for vocabulary scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	.352	.176	.338	.716
Errors	25	13.006	.520		
<b>Within Groups</b>					
Time	1	.002	.002	.008	.929
Time x Type	2	.728	.364	1.726	.199
Errors	25	5.272	.211		



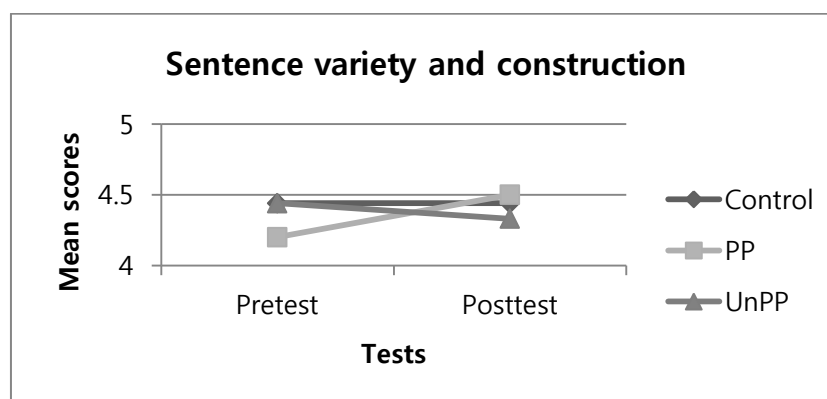
**Figure 4.4 Vocabulary mean scores across group**

In the same manner, the results of repeated measures ANOVA also indicated that planning had no statistically significant effect on the vocabulary scores between groups  $F = .338, p > .05$ . Also, there was no significant Time x Group interaction,  $F = 1.726, p > .05$ .

When examining the mean scores among groups, as with the development of ideas section, the Control and the UnPP groups had the same pattern with a slight decrease while PP group had a trivial increase.

**Table 4. 14 ANOVA for sentence variety and construction scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	.85	.042	.044	.957
Errors	25	24.272	.971		
<b>Within Groups</b>					
Time	1	.055	.055	.308	.584
Time x Type	2	.434	.217	1.207	.316
Errors	25	4.494	.180		



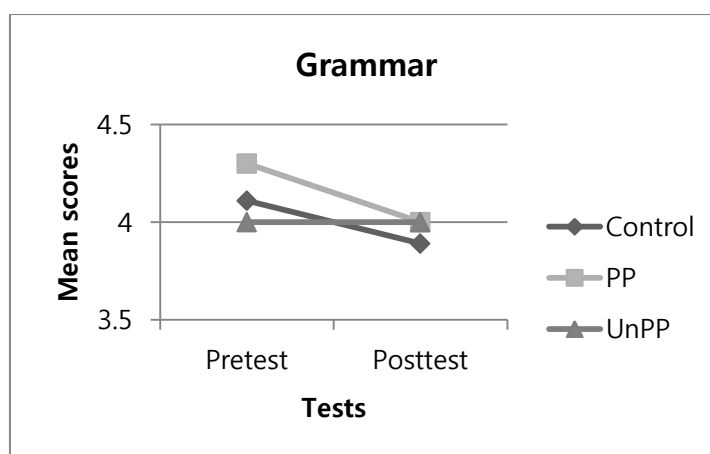
**Figure 4.5 Sentence variety and construction mean scores across groups**

As with other criteria mentioned earlier, sentence variety and construction showed similar results in Table 4.14. The results of repeated measures ANOVA were not statistically significant,  $F = .044$ ,  $p < .05$ . Furthermore, there was no significant Time  $\times$  Type interaction,  $F = 1.207$ ,  $p > .05$ .

As shown from Figure 4.5, the control and the UnPP groups had no change or a slight decrease while the PP group rose. Also, the mean score band was narrow, demonstrating a minimal difference between the pretest and posttest.

**Table 4. 15 ANOVA for grammar scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	.289	.145	.180	.836
Errors	25	20.050	.802		
<b>Within Groups</b>					
Time	1	.423	.423	1.815	.190
Time x Type	2	.226	.113	.484	.622
Errors	25	5.828	.233		

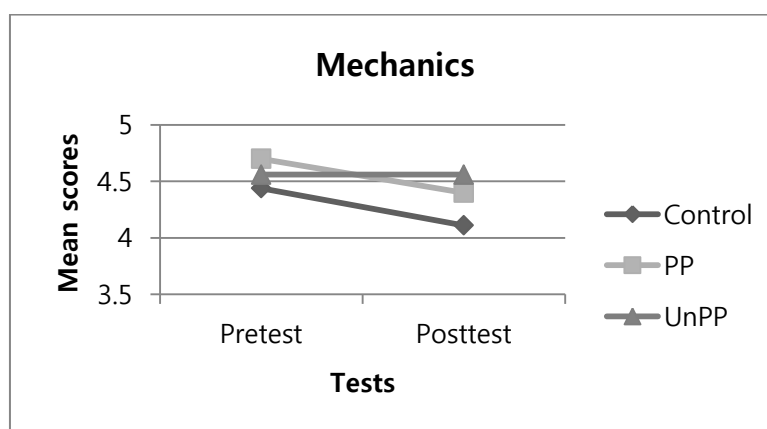


**Figure 4.6 Grammar mean scores across groups**

The results for the grammar concurred with the previous findings. Planning had no statistically significant effect on the grammar as well,  $F=.180$ ,  $p>.05$ . There was not any significant Time  $\times$  Type interaction,  $F=1.815$ ,  $p>.05$ , either. As shown from Figure 4.6, the mean values had a decrease in control and thePP group. On the other hand, the UnPP group stayed the same across the tests.

**Table 4. 16 ANOVA for mechanics scores of writing tests**

Source	df	SS	MS	<i>F</i>	<i>p</i>
<b>Between Groups</b>					
Type	2	.923	.462	.444	.647
Errors	25	26.006	1.040		
<b>Within Groups</b>					
Time	1	.622	.622	1.548	.225
Time $\times$ Type	2	.307	.154	.382	.686
Errors	25	10.050	.402		



**Figure 4.7 Mechanic mean scores across groups**

The results for the final criterion, mechanics, corresponded with the previous reports. There were no statistically significant differences between

groups,  $F = .444$ ,  $p > .05$ . There was no significant Time  $\times$  Type interaction,  $F = 1.548$ ,  $p > .05$ .

The mean values also conformed to other criteria showing declines and no change. The mean scores of the control and the PP groups dropped while those of the UnPP group remained unchanged.

Overall, the results of the series of repeated measures ANOVA showed that planning had no significant effects on the writing in any of the analytic scores, nor in composite score. Moreover, there was no significant Time  $\times$  Type interaction. Even though there were slight increase and decrease patterns across groups, there were no substantial differences in the scores.

#### **4.5. Analysis of Covariance (ANCOVA)**

##### **4.5.1 Proficiency mediating influence on effects of planning conditions on writing scores**

In order to examine whether test-takers' proficiency level have any mediating effect on impacts of planning conditions on writing scores, a series of 2 (group)  $\times$  3 (planning conditions) mixed model analyses of covariance (ANCOVA) were conducted.

Analytic and composite pre-test scores were controlled as covariate. That is, preexisting differences between two proficiency groups were adjusted. In this way, the two groups scored differently on the pre-test, and the group mean scores on the post-test are adjusted to account for the difference on the pre-test.

**Table 4. 17. ANCOVA for Proficiency mediating effect**

	<b>Source</b>	<b>df</b>	<b>SS</b>	<b>MS</b>	<b>F</b>	<b>p</b>
DI	Proficiency	1	.082	.082	.076	.786
	Type	2	.505	.253	.235	.792
	Proficiency x Type	2	.386	.193	.180	.837
	Errors	21	22.536	1.073		
O	Proficiency	1	.704	.704	.925	.347
	Type	2	.253	.126	.166	.848
	Proficiency x Type	2	.250	.125	.164	.850
	Errors	21	15.987	.761		
V	Proficiency	1	.501	.501	1.467	.239
	Type	2	1.319	.660	1.932	.170
	Proficiency x Type	2	.768	.384	1.124	.344
	Errors	21	7.172	.342		
SV	Proficiency	1	.179	.179	.747	.397
	Type	2	.350	.175	.729	.494
	Proficiency x Type	2	.110	.055	.229	.797
	Errors	21	5.037	.240		
G	Proficiency	1	.003	.003	.006	.939
	Type	2	.157	.078	.183	.834
	Proficiency x Type	2	.014	.007	.017	.983
	Errors	21	8.977	.427		
M	Proficiency	1	.045	.045	.057	.813
	Type	2	.678	.339	.434	.653
	Proficiency x Type	2	1.909	.954	1.224	.314
	Errors	21	16.375	.780		
Composite	Proficiency	1	3.290	3.290	.383	.542
	Type	2	11.170	5.585	.651	.532
	Proficiency x Type	2	2.075	1.038	.121	.887
	Errors	21	180.147	8.578		

As shown in Table 4. 17, the results indicated that test-takers' proficiency level turned out to have no statistically significant effects on analytic writing scores under different planning conditions. These results were also consistent with the composite scores. In other words, Korean EFL test-takers' proficiency had no mediating effect on writing performance depending on planning conditions.

Consequently, the result of this study indicates that the Korean EFL learners' proficiency level had little effect upon the writing performance,

showing no significant difference nor interaction effect between Proficiency x Type in any of the six analytic criteria and even composite scores on writing assessment tasks under pressured and unpressured within-task planning conditions.

#### **4.5 Rater's Post-Rating Feedback**

In order to supplement the quantitative results, the raters provided post-rating feedback by responding to a short questionnaire about writing tests, analytic scale, general impressions and comments on the writing samples (see Appendix 7).

When it comes to the overall opinion about the writing samples, both raters agreed on the fact that participants' writing ability seemed to be above intermediate level. Overall, all participants wrote quite well. Also, they agreed that the writing prompts were general topics which were not difficult to write and were clear to understand.

However, they responded that the most difficult part when rating writing samples was the scale whose descriptors were too general, not specific and not clear containing some abstract explanations. It was not easy to interpret and understand some descriptors. Even though the scale included essential sub-categories of writing such as development of ideas, organization, vocabulary, sentence variety and construction, grammar and mechanics with the score range from 1 to 5 or 6, there were no cases in which scores of 1 or 2 points were assigned to writing samples. Therefore, raters said that the score band was not very appropriate for learners.



The mechanics section was the easiest to rate since it was readily observable within writing texts. Sentence variety and construction and vocabulary sections followed as they were also easily identified. On the other hand, the development of ideas and organization sections were the trickiest parts to grade as raters should read logically and examine details carefully. One rater said that the scores of 3 or 4 were the hardest to discern.

Moreover, one rater suggested that it is necessary to develop more appropriate analytic scales for learners' writing and language proficiency level. In the case of grammar, writers could get 3 points though they made errors on 70% of their writing.

#### **4.6 Questionnaires and Interviews**

All participants filled in two types of questionnaires: one was a pre-test questionnaire about their writing strategies and the other was a post-test questionnaire about the writing tests they took. The latter was mainly designed to collect participants' feedback regarding writing tests, planning time and behavior, and their perceptions about planning. All responses were collected and tallied. Some important results of the tally are shown graphically in Appendix 8. Furthermore, some participants from all groups were invited to attend a brief individual interview based on their responses to the questionnaires. The interview was conducted immediately after the post-test questionnaire.

The pre-test questionnaire consists of four sections: general English learning strategies; general English writing strategies; before-writing

strategies; and during-writing strategies. Each section has 7 questions. Question 1 from section 1 is about whether they find as many ways as they can to use English. Fifty-five percent of the control group, 60% of the PP group and 55% of the UnPP group agreed that they try hard to use English which can indicate that they are highly motivated to learn English.

In terms of the ways of using English, they had quite a lot of opportunities to read English since 44% of the control group, 80% of the PP group and 50% of the UnPP group strongly agreed on question 2. On the other hand, only 11% of the control group, 10% of the PP group and 22% of the UnPP group responded that they have many chances to write. It seemed that they have relatively few opportunities to write in English. Even though their majors varied, some of the participants interviewed said that they read various textbooks in English, but they do not have to write in English. One of the participants said that she wanted to practice English writing, but it was difficult to get feedback on her English writing.

According to Questions 8 and 9 from the general English writing strategies section, 44% of the control group, 40% of the PP group and 55% of the UnPP group participants strongly agreed that they often write in Korean. However, one participant each from the PP and UnPP groups each agreed on question 9 that they often write in English. Rather, 66% of the control group, 50% of the PP group and 77% of the UnPP group strongly disagreed that they write in English. It can be inferred from this that the participants are not exposed to writing in English a lot.

What is noticeable is that the participants were well aware of some writing

skills and made use of them. In the case of Question 12, 50% of the control group, 90% of the PP group and 88% of the UnPP group participants strongly agreed that they use cohesive devices in English writing. Moreover, 88% of the control group, 90% of the PP group and 77% of the UnPP group responded to question 26 that they do paraphrasing.

Regarding before-writing strategies, which this study mainly focused on, 66% of the control group, 50% of the PP group and 66% of the UnPP group members answered that they do brainstorming. Also, participants were quite familiar with making plans since 66% of the control, 10% of the PP and 55% of the UnPP groups responded to question 17 that they make plans and outlines in Korean before writing in English. On the contrary, 60% of PP group and 11% of the UnPP group replied to question 18 that they make plans and outlines in English. The participants interviewed indicated that it is too great a burden for them to make plans in English since they cannot proceed with thinking processes in English (see Appendix 8).

When it comes to Questions 23 and 24, 22% of the control group, 30% of the PP group and 33% of the UnPP group strongly agreed that they kept the planning made before writing. On the other hand, all participants responded that they try to change or make ideas clearer while they are writing. Interviewees said that recently many English writing tests are conducted by computer, so it becomes easy for test-takers to cut, or copy and paste what they write. This seems to affect their writing routines.

Table 4.18 presents post-test questionnaire data among the groups. Ninety percent of participants from both the control and PP group, and 55% of

subjects from the UnPP group disagreed on the fact that the testing time was short. Since all of them had 30 minutes for writing tests in total, it appears to be quite enough and sufficient time for them to write essays. Ten percent of the PP group responded that they needed more planning time as they were provided only 5 minutes to plan while 45% of the UnPP group also answered that they needed more time even though they had full control of managing planning time.

**Table 4.18 Post-questionnaire data**

		Strongly disagree		Strongly agree		
Question	group	1	2	3	4	5
Testing time was short	Control	4(45%)	4(45%)	1(10%)	0	0
	PP	2(20%)	7(70%)	1(10%)	0	0
	UnPP	0	5(55%)	2(22%)	2(22%)	0
Need more planning time	PP	1(10%)	3(30%)	5(50%)	1(10%)	0
	UnPP	0	3(33%)	2(22%)	4(45%)	0
Appropriate planning time	PP	0	4(40%)	3(30%)	3(30%)	0
	UnPP	1(10%)	4(45%)	2(22%)	2(22%)	0
Tests were difficult	Control	4(45%)	5(55%)	0	0	0
	PP	2(20%)	6(20%)	2(20%)	0	0
	UnPP	0	7(77%)	2(22%)	0	0
Show English writing skills	Control	0	1(10%)	4(45%)	3(33%)	1(10%)
	PP	0	1(10%)	4(40%)	4(40%)	1(10%)
	UnPP	0	0	3(33%)	6(66%)	0
		Prompt A	Prompt B	Neither		
More difficult topic	Control	0	5(55%)	4(45%)		
	PP	5(50%)	3(30%)	2(20%)		
	UnPP	3(33%)	3(33%)	3(33%)		

When it comes to the appropriate planning time, 40% of the PP group responded that three to five minutes is enough while 45% of the UnPP group

did. It is interesting that 30% of PP as well as 22% of the UnPP group replied that seven to ten minutes is needed for planning. Actually, there were 6 participants out of 9 who spent more than 5 minutes for planning in the UnPP group.

Ninety-nine percent, 40% and 77% of each group disagreed that the tests were difficult. This means that they didn't think either the pretest or posttest were difficult. Forty-three percent of the control group, 50% of the PP group and 66% of the UnPP group agreed on the fact that the tests showed their English writing skills well.

Fifty-five percent of the control group responded that prompt B, which was about the best ways to relieve stress was more difficult than prompt A, which asked opinions about banning smoking in public areas. However, 55% of the control group also answered that neither of the topics was difficult. On the contrary, 50% of the PP group said prompt A was more difficult than Prompt B. The responses of the UnPP group had an even distribution across the difficulty of prompts: 33% for prompt A; 33 % for prompt B; and 33% for neither of the topics being difficult. According to the follow up interviews, participants said that prompt A was difficult in that they could not come up with good supporting ideas while for prompt B, it was hard to choose the best ways out of many methods to relieve stress, which required expressing more individual preference with good supporting details.

The results of interviews were consistent with their responses to questionnaires and supplied good supplementary information. Seventy-five percent of the participants believed that planning is important regardless of

the test modules: either computer-based or paper-based. They tended to consider planning helpful to organize key concepts and ideas. Some replied that planning enhances writing fluency as they can write more and planning makes their writing rich in content. However, they were likely to make plans simple. Once they decided to take a position, they attempted to think about the reasons why they took the position in argumentative writing. They usually tried to form a structure of the whole composition using simple words. Furthermore, they reported that planning benefits them to keep cohesion of writing texts.

Nevertheless, there were some participants (25%) who answered that they did not have to make plans as they revised the texts simultaneously as they were writing. Since the test was conducted by computer, they thought they could incorporate revisions into the text easily and save time. Moreover, a few learners answered that it is necessary to make plans only if the topic or genre is difficult. They felt that they do not have to make plans for simple description or narrative writing tasks.

#### **4.7 Analysis of Korean EFL test-takers' writing plan behavior**

Participants' planning sheets from the PP and the UnPP group were collected and analyzed according to their proficiency levels. There were 19 planning sheets from 10 advanced learners and 9 from intermediate learners.

With regard to the language they used for planning, 66% of the advanced group used English while 40% of the intermediate group did. One learner from the advanced group used only Korean while 40% of the intermediate

group did. Twenty-two percent of the advanced group and 20% of the intermediate group used both Korean and English, especially using English for key words.

According to the results of the UnPP group, the average time spent on planning was 5.06 minutes. Even though the number of participants was small to compare, advanced learners spent 6.27 minutes on planning while intermediate learners used 4.08 minutes. It is likely that advanced writers spend more time on planning than intermediate writers.

Regardless of proficiency, learners usually wrote down key words and supporting ideas for their position. Results indicate that the learners first took a position about the topic, and then tried to come up with two or three supporting details for the position. However, they did not attempt to consider the opposite position offering a rebuttal or argument.

All learners seemed to be familiar with making plans by organizing a structure consisting of introduction, body and conclusion. Even 60% of the intermediate group applied this structure as well. It should also be considered that 6 participants had already taken lessons for TOEFL essays. Moreover, as seen from the questionnaire, learners used cohesive devices frequently marking different paragraphs in planning sheets. Some learners listed reasons and examples without an orderly manner.

Over 80% of learners usually used simple words and made plans short while a few learners wrote full sentences in English as if drafting. In other words, the Korean EFL learners who joined this study mainly used organized notes, which can be equated to outlining.

## **CHAPTER V**

### **DISCUSSION**

This chapter discusses the major findings with respect to the research questions posed for the study earlier in Chapter 1: The effects of within-task planning conditions in terms of each analytic score and composite score of writing assessment task; the mediating effect of test-takers' proficiency on the impacts of planning conditions on writing scores; and planning patterns and behaviors of Korean EFL test-takers.

#### **5.1. Effects of within-task planning conditions on the analytic and composite Scores**

The statistical results indicated that the effects of pressured and unpressured within-task planning conditions on argumentative writing had no significant influence on the composite scores as well as analytic scores of participants. When the planning effects were analyzed through raw scores, there were slight increases and decreases among groups. However, the results of a series of the repeated measures ANOVA demonstrated that this difference was not statistically significant. Also, there was no Time  $\times$  Type interaction, which meant that there was no significant variance between pretest and posttest among differently conditioned groups.

In regard to the descriptive statistics, the PP group had a slight increase in composite scores and some of analytic scores: development of ideas; organization; vocabulary; and sentence variety and construction sections.



Since the former two sections, which were development of ideas, and organization, can belong to sub-skills of writing competence about how to organize and develop paragraphs, it is likely that the PP group benefited slightly in writing skills. Also, the latter two sections, which were the vocabulary, and sentence variety and construction sections, represent language skills. However, it is not obvious that the PP group also benefited slightly in language competence since the average TEPS score of the PP group was higher than that of the other two groups, which was not significant.

On the other hand, the grammar and mechanics sections scores declined slightly, which somewhat echoes the reports by Hulstijn (1984). Hulstijn suggested that giving test-takers time to plan on-line and monitor their output appears to have a clear impact on accuracy; however, if they use the time to plan content, no effect on accuracy is observed.

In case of the control and the UnPP groups, they showed similar downward patterns across almost all analytic and composite scores. Even though the control group was controlled to attempt to prevent them from planning before writing, post-test questionnaire and interviews revealed that some of participants still tried to make plans, since they could imagine some key concepts about the writing prompt in their mind; this can be equal to one of planning strategies: mind mapping. Planning is considered to be the ability to abstract ideas meta-cognitively about one's relevant linguistic knowledge, and to be able to retain at least some of this plan and subsequently act on it (Bastone, 2003). In this regard, participants in this study already had some planning abilities. Moreover, it appeared from the questionnaires that 75% of

the participants believed that planning is important and useful, which motivates them to keep utilizing planning in their performance.

What is noteworthy is that the UnPP group had a slight increase in the organization section, which was not found in the control group. As the UnPP group was told to plan before writing at their own pace, they seemed to have an opportunity to plan carefully and effectively. Effective planning is a skillful and demanding activity requiring a careful, conscious, and selective engagement with language (Skehan, 1998; O'Malley & Chamot, 1990). In this respect, within-task planning has a slight effect on organization which is more related to writing skill. In other words, despite small raw values, learners benefited from within-task planning, but the pressured condition had more influence on learners' writings than the unpressured condition in term of language competence and writing skills.

The present study examined the effects of planning on the writing performance in a testing situation. If planning time can affect a test-taker's performance, then it ought to be considered when designing writing tests. However, as seen from the previous chapter, there were no statistically significant differences among the groups. This result echoes the findings from Wigglesworth (1997), even though that studied looked at oral performance. Wigglesworth reported that at least for some learners and in some tasks, planning time can help to improve the performance of test-takers, but that this effect is not evident in external rating. Since this study also rated participants' writing performance by discourse analytic measures, the same results shown gave more insights into testing situations. It is possible, then, that the testing

context constrains the beneficial effects of planning. This suggested that the psychological context of a task constitutes an important dimension that needs to be taken into account in planning studies (Bastone, 2003).

## **5.2 Korean EFL test-takers' proficiency mediating influence on effects of planning conditions on writing scores**

The analyses of the analytic and the sum of analytic scores according to test-takers' different proficiency concurred with the results on the first research question. The statistical results indicated that Korean EFL test takers' different levels of proficiency had no significant influence on the composite scores as well as analytic scores under different within-task planning conditions on argumentative writing. That is, the proficiency level does not mediate the effects of planning conditions on writing scores. However, since there were a small number of participants in this study, the values which can represent the effect of proficiency and planning should be interpreted carefully.

It is possible that in the previous study, different proficiency groups could have different tendency and advantages of planning, which could mediate the effects of planning conditions. Wigglesworth (1997) suggested that the advanced learners may focus on the form and complexity of their linguistic output, while low-proficiency learners may focus on content. Also, it was claimed that for the advanced learners, planning time may be beneficial in a situation when the cognitive load becomes a big burden. Consequently, in the more difficult tasks, it is possible that planning time plays a significant role in reducing the load. With simple tasks, there is no need to have planning time

because test-takers possess sufficient cognitive capacity to manage the task and resources. In fact, for the high proficiency learners, planning time does not make a big difference on the easier tasks, but where the task is more difficult and the learners are running out of cognitive resources, some differences emerge. However, this difference is not pronounced in low proficiency group. Low-level learners do not appear to be advantaged by planning time.

For a more complete study, future research should include more learners that can show some trends or any mediating effects from different levels of proficiency.

### **5.3. Planning Patterns and Behaviors of Korean EFL Test-takers**

Participants' responses to questionnaires and interview provide valuable information about planning patterns and behaviors of Korean EFL learners.

First of all, 75% of the participants believed that planning is important regardless of the test modules: either computer-based or paper-based. They tended to consider planning helpful to organize key concepts and ideas.

In terms of using English, they had quite a lot of opportunities to read English while they had few opportunities to write. Soon after the placement test, some of the participants said that it was not familiar for them to write even some sentences describing pictures in English.

However, they often write in Korean. In this respect, the language they used for planning varied from Korean to English according to their proficiency. More than half of the participants from the advanced group used English

while few of those from the intermediate group did so. One learner from the advanced group used only Korean, while almost half of the learners from the intermediate group did. Twenty-two percent of the advanced group and 20% of the intermediate group used both Korean and English, especially using English for key words. The participants interviewed indicated that it is too much a burden for them to make plans in English, since they cannot proceed with thinking processes in English.

Regarding before-writing strategies, which were the main focus of this study, more than 60 % of the participants answered that they do brainstorming. With respect to the method of planning, over 80% of learners usually used simple words making plans short so that they are essentially equivalent to outlining, while a few learners wrote full sentences in English, like drafting. Some also made lists of examples and key concepts. Korean EFL learners seemed to use simple way to make plans.

As raters reported, the participants' writing samples were above the intermediate level and they responded that they already got used to some writing skills: using cohesive devices; and paraphrasing. This also showed that participants had extensive vocabulary knowledge.

When it comes to appropriate planning time, almost half of the participants agreed that three to five minutes is enough while there were some learners who felt that seven to ten minutes is needed for planning. Consistent with the responses, the average time spent on planning was 5.06 minutes according to the results of the UnPP group. Even though the number of participants was small for comparison purposes, advanced learners spent 6.27 minutes on

planning while intermediate learners did 4.08 minutes. It is likely that advanced writers spend more time on planning than intermediate writers. This result displayed the same tendency as the previous studies, in that low-proficiency writers plan less, while high-proficiency writers plan more (Raimes, 1985, 1987; Torrance, 1996; Zamel, 1982, 1983).

## **CHAPTER V**

### **CONCLUSION**

#### **6.1. Conclusions and implications**

The results of this study indicated that within-task planning conditions had no statistically significant effects on Korean EFL learners' performance on argumentative writing tasks. The writing task scores for the composite as well as all six analytic criteria showed little indication of differences in performance across the pretest and posttest. Moreover, within-task conditions had no statistically significant effects on different proficiency levels of Korean EFL learners' writing production. Contrary to studies that have found some beneficial and detrimental impacts on the planning of writing tasks, this study found that the writing scores were not affected by the planning conditions at a statistically significant level.

Based on the findings of this study, one major implication is the assignment of within-task planning time. Since there were no significant effects of three experimental planning conditions, the current large-scale standardized writing tests format which limits writing time incorporating all the processes of writing can remain unchanged. Prewriting activities might have some beneficial effects on writing performance. However, what this study does indicate is that in a testing situation, a pressured or unpressured planning condition seemed not to affect writing performance in a significant way. This empirical evidence also provides support for existing writing test formats which do not give planning time separately.

However, despite small differences, there were slightly positive impacts on the composite scores as well as analytic ones. Furthermore, participants' responses to questionnaires and interviews revealed that the participants consider planning to be an effective and useful activity to improve their writing quality and thus, they have already attempted to utilize planning while they are writing. In this regard, assigning within-task planning time can be taken into account as an alternative to give learners more opportunities to show their writing ability as well as linguistic knowledge. When it comes to recent strategy-based instruction, planning can also benefit students in terms of practicality and authenticity. It is also interesting that making plans in L1 can also have a positive effect on L2 writing. The positive transfer from L1 to L2 was identified by the questionnaires and interview in this study. The instructions and training about making plans and outlines Korean EFL learners have received help them write argumentative texts in English.

Even though more research is needed to implement the planning stage in the writing tests, within-task planning can be considered as a viable component of the writing task for future English writing assessment in the Korean EFL context.

## **6.2 Limitations and future studies**

There were a number of limitations of this study in terms of the methodology and analyses of data collected. First of all, this study was based on a small sample size ( $n=28$ ). Because of the small number of participants assigned in three experimental conditions, the findings should be interpreted with caution.



Therefore, a larger sample size would provide more accurate analysis of results. It is also assumed that a larger sample size of different levels of proficiency learners could give more insights on the effects of planning.

Moreover, there was a posttest which measured the effects of three experimental conditions by only one task. Several tasks could provide a more accurate investigation of planning effects on writing scores. Even though the order of prompts was counterbalanced to control the difficulty difference, the responses to questionnaires suggested that there may have been some individual variation in prompts difficulty. In other words, some participants thought neither prompts was difficult, while others considered one or another prompt to be more difficult. The prompt difficulty should be more carefully arranged for the future study.

Most importantly, a finer-grained analytic rubric should be developed and applied to the rating of writing samples. According to the raters' post-rating feedback, it is necessary to improve the scoring rubrics by identifying and specifying the elements of each component of writing that are currently not captured by the scoring descriptors. The score band should also be scrutinized to provide more accurate writing scores in future investigation. Research oriented to better test design should, for practical reasons, rely on scores derived from accurate rating.

This study attempted to examine the planning strategies of the learners by using a pre-writing questionnaire, however, it would have been better if the questionnaire had been administered immediately after learners took the experimental writing tests dealing with what strategies they also used during

the tests. Moreover, the question prompts should have narrowed down the range of writing, limiting it to only in English to offer more implications of Korean EFL learners' writing behavior. Future studies can also consider using a different method like a think-aloud technique in order to trace learners' cognitive processes to examine the planning patterns of learners further.

Despite such limitations, the significance of the study is that it attempted to investigate the effects of within-task planning conditions on argumentative writing assessment tasks. By applying an analytic rubric to assess writing performance and examine planning behavior of Korean EFL test takers, this study helps reach a deeper understanding of how Korean EFL test-takers utilize planning in their writing.

## References

- Bachman L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University press
- Bachman, L.F., Davidson, F., Ryan, K. and Choi, I-C. (1995). *An investigation into comparability of two tests of English as a foreign language*. The Cambridge-TOEFL comparability study. Cambridge: Cambridge University Press.
- Bastone, R. (2002). *Planning as discourse activity: A sociocognitive view*. Planning and task performance in a second language. John Benjamins B.V.
- Bland J.M., & Altman D.G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, 314-572.
- Chai, C. (2006). Writing plan quality: Relevance to writing scores, *Assessing Writing*, 11, 198-223.
- Chieko Kawauchi (2005). *The Effects of strategic planning on the oral narratives of learners with low and high intermediated L2 proficiency*. In R. Ellis (Ed.) *Planning and task performance in a second language*. Amsterdam: John Benjamins: 143-164.
- Cumming, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics*, 18, 61-78.
- Dellerman, P. Coirier, P., Marchand, E. (1996). Planning and expertise in argumentative composition. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.). *Theories, models, and methodology in writing research* (pp 182-195). Amsterdam: Amsterdam University Press.
- Elder, C. & Iwashita, N. (2005). *Planning for test performance*. In R. Ellis (Ed.)

Planning and task performance in a second language. Amsterdam: John Benjamins: 219-238.

Ellis, R., & Yuan, F. (2004a). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59-84.

Ellis, R. (2004b). *Planning and task performance in a second language*. John Benjamins B.V.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323.

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing*. London: Longman.

Glynn, S. M., Britton, B., Muth, D., & Dogan, N. (1982). Writing and revising persuasive documents: Cognitive demands. *Journal of Educational Psychology*, 74, 557–567.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ:Lawrence Erlbaum.

Hayes, J. and J. Gradwohl Nash. (1996). *On the nature of planning in writing*. In C. Levy and S. Ransdell (eds), *The Science of Writing*. Mahwah NJ: Lawrence Erlbaum

Hinkel, E. (2002). *Second Language Writers' Text: linguistic and rhetorical features*. Lawrence Erlbaum Associates, Inc

Hulstijn, J. and W. Hulstijn. (1984). Grammatical errors as a function of

- processing constraints and and explicit knowledge. *Language Learning*, 34, 23-43.
- IBM Corporation. (1989, 2011). IBM SPSS Statistics 20 Core System User's Guide.
- Jacobs, S. (1982). *Composing and coherence: The writing of eleven pre-medical students*. Linguistics and Literacy Series 3. Washington, D.C.:Center for Applied Linguistics.
- Kroll, B.(1990). *Second Language Writing: Research insights for the classroom*. Cambridge applied linguistics series.
- Lee, Y. W., Gentile C, and Kantor R (2008) Analytic Scoring of TOEFL CBT Essays: Scores From Humans and E-rater, ETS *Research Report (No 81)*, Princeton, NJ: Educational Testing Service
- McCutchen, D. (2011). From Novice to Expert: Implications of Language Skills and Writing-Relevant Knowledge for Memory during the Development of Writing Skill. *Journal of Writing Research*, 3(1), 51-68.
- Mark D. J., Leonardi M., & Anthony A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, 21, 264-28
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83–108.
- Nakakubo, Takako. The effects of planning on second language oral performance in Japanese: processes and production." doctoral dissertation, University of Iowa, 2011.

- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109–148.
- Paltridge, B. (2004). Academic writing. *Language Teaching*, 37, 87–105.
- Powers, D. (2010) The Case for a Comprehensive, Four Skills Assessment of English Language Proficiency. *R&D Connections*, No. 14. Princeton, NJ: Educational Testing Service
- Purpura, J. E. (2004). *Assessing Grammar*. Cambridge University Press
- Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19, 229–258.
- Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning*, 37, 439–468.
- Regina, L. S., Mary, K. R., Joann, R.K. (2001). *Refining composition skills: rhetoric and grammar*. Heinle & Heinle
- Reinard, J.C.(2006). *Communication research statistics*. Thousand Oaks, CA: Sage.
- Roca de Larios, J., Marin, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, 51, 497–538.
- Roca de Larios, J. R., Murphy, L., & Marin, J. (2002). Critical examination of L2 writing process research. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 11–48). Dordrecht, The Netherlands: Kluwer Academic.
- Sasaki, M. (1996) *Second language proficiency, foreign language aptitude,*

- and intelligence: *Quantitative and qualitative analysis*. New York: Peter Lang.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9, 259-291.
- Sawaki, Y., Sticker, L., & Oranje, A. (2008) *Factor structure of the TOEFL® internetbased test (iBT): Exploration in a field trial sample* (ETS Research REP. NO. RR-08-09) Princeton, NJ: Educational Testing Service.
- Shin, Y-S (2008). *The effects of planning on L2 writing: a study of Korean learners of English as a foreign language*. Doctoral dissertation. University of Iowa.
- Shin, S-K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, 22(1), 31-57.
- Silva, Tony. (1993). Towards an understanding of the distinct nature of L2 writing: The ESL research and implications. *TESOL Quarterly* 27: 657-77.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1, 185–211.
- Skehan, P. (1998). *A cognitive approach to language learning*. New York: Oxford University Press.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.
- Tavakoli, P. and Skehan, P. (2005). *Strategic planning, task structure and performance testing*. In R. Ellis (Ed.) *Planning and task performance in a*

- second language. Amsterdam: John Benjamins: 239-277.
- Torrance, M. (1996). Is writing expertise like other kinds of expertise? In G. Rijlaarsdam, H. v. d. Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 268–281). Amsterdam: Amsterdam University Press.
- Vogt, W.P. (2005). *Dictionary of Statistics & methodology: A nontechnical guide for the social sciences* (3 rd ed.). Thousand Oaks, CA: Sage.
- Wang, W., & Wen, Q. (2002). L1 use in the L2 composing processes: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11, 225–246.
- Weigle, S.C.(2002). *Assessing Writing*. Cambridge University Press
- Wendel, J. N. (1997). *Planning and second-language narrative production*, Temple University, Philadelphia, PA.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14, 85-106.
- Zamel, V. (1982). Writing: The process of discovering meaning. *TESOL Quarterly*, 16, 195–209.
- Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, 17, 165–187.



## Appendix 1

### Participant's Background Information

Please complete the following;

1. Name : \_\_\_\_\_ Age \_\_\_\_\_ Major \_\_\_\_\_
2. Time spent abroad in an English speaking country: 영어 사용권 나라 거주경험  
\_\_\_\_\_ (months or years)
3. Your TEPS Score(s): \_\_\_\_\_ (in total) 텡스 총점  
☞ Grammar section \_\_\_\_\_/100 ☞ Vocabulary section \_\_\_\_\_/100
4. What was your score in the writing section (if any) ? \_\_\_\_\_
5. Have you ever taken an English writing class or instruction? ( Yes / No )  
영어글쓰기 관련 수업수강여부
  - 5.1 How long did you take lessons? \_\_\_\_\_
  - 5.2 What was the class mainly about? \_\_\_\_\_
6. If you have difficulty in writing, what do you think your problem area is?  
영어 글쓰기가 어렵다면 어느 영역이 가장 고민입니까?
  - (a) grammar
  - (b) vocabulary
  - (c) organization 구성
  - (d) content 내용
  - (e) mechanics (punctuation, spelling, capitalization, and indentation )  
마침표 등의 문장부호, 철자, 대문자, 띄어쓰기 등
7. Rate your own English ability on the following scale by circling the number.

	minimal			near-native	
<b>Grammar</b>	1	2	3	4	5
<b>Vocabulary</b>	1	2	3	4	5
<b>Organization</b>	1	2	3	4	5
<b>Content</b>	1	2	3	4	5
<b>Mechanics</b>	1	2	3	4	5

## Consent Form

### **Purpose of study:**

You are invited to participate in a research study about an English writing test. This study aims to learn about Korean EFL learners' writing and their writing strategies. Your writing will be analyzed in order to see how much it changes in different conditions.

### **Procedure and Time required:**

The experiment will be divided into two parts and will take 2 days to complete it. On the first day, you will fill out the background information sheet and consent form. Then, you will take a placement test for 15 minutes. You will also fill out a questionnaire about writing strategies. On the second day, you will take two main writing tests and each test will take 30 minutes. After finishing two writing tests, you will fill out the post-questionnaire. Finally, the researcher is going to conduct a short oral interview based on your responses to the questionnaires.

### **Access to existing records:**

You were requested to provide me with your TEPS scores.

### **Compensation:**

You will receive a payment of 20,000 won when you have finished taking all tests, questionnaires and an interview. You can also get the feedback and score of your writing tests if you want.

### **Confidentiality:**

Your identity will be kept confidential as your name or any other information that could possibly indicate your identity will be excluded from the final report of this research study.

### **Voluntary participation:**

Your participation in this study is completely voluntary. If you chose not to participate in this study, this will have no effect on the services or benefits you are currently receiving. You may choose to stop participating in the study at any time. This will have no effect on your current or future relation with Seoul National University.

### **Contact information:**

If you have any questions, you can contact the researcher at;

Ms. Junghyun Park                      010 6363 2196      email [pocari32@snu.ac.kr](mailto:pocari32@snu.ac.kr)

I have read and understood the information stated above and consent to participate in this study.

Participant;

Name \_\_\_\_\_

Signature: \_\_\_\_\_

Researcher

Name      Junghyun      Park

Signature: \_\_\_\_\_

Date \_\_\_\_\_

## Appendix 2

### Self-assessment of English Writing Strategies

Please read each statement below, then fill in the circle under the number that most accurately reflects your opinion or attitude.

다음에 나오는 의견에 부합하는 숫자나 상자에 체크해 주세요.

#### Section A (1~7)

**1. I try to find as many ways as I can to use my English. \***

나는 영어를 사용할 수 있는 가능한 많은 방법을 시도한다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**2. I look for opportunities to read as much as possible in English. \***

영어를 읽을 수 있는 기회를 가능한 많이 찾는다. (영어 읽기 기회가 많다.)

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**3. I look for opportunities to speak as much as possible in English. \***

영어를 말할 수 있는 기회를 가능한 많이 찾는다. (영어로 말할 기회가 많다.)

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**4. I look for opportunities to listen as much as possible to English. \***

영어를 들을 수 있는 기회를 가능한 많이 찾는다. (영어 듣기 기회가 많다.)

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**5. I look for opportunities to write as much as possible in English. \***

영어로 글을 쓸 수 있는 기회를 가능한 많이 찾는다. (영어로 글쓰기 기회가 많다.)

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**6. I notice I am tense and nervous when I use English. \***

영어를 사용할 때 긴장되고 두렵다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**7. I notice my mistakes and use that information to help me study English.\***

내가 범한 오류와 실수를 인지하여 영어 학습을 위해 활용한다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**Section B (8~14)**

**8. I often write in Korean.\***

나는 한국어로 자주 글을 쓴다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**9. I often write in English.\***

나는 영어로 자주 글을 쓴다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**10. I often read English texts such as books, magazines, or articles.\***

나는 책이나 잡지, 기사와 같은 영어로 씌여진 글을 자주 읽는다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**11. I use English words I know in different ways.\***

나는 내가 알고 있는 영어 단어를 다양한 방식으로 활용한다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**12. I use cohesive devices in my English writing. (e.g. thus, moreover, however, and so on.)\***

나는 영어글쓰기에서 접속사와 같이 단락을 연결해주는 단어를 사용한다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**13. I express a particular meaning in different grammatical forms.\***

나는 특정 의미를 다양한 문법 형식으로 표현할 수 있다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**14. I use appropriate word order and grammatical patterns. (e.g. tense, agreement, and pluralization)\***

나는 적절한 어순과 문법 패턴들을 사용한다. (시제, 일치, 복수형 만들기 등)

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

## Section C (15~21)

**15. I consider the task or assignment and instructions carefully before writing. \***

나는 글쓰기 전 글쓰기 시험과 문제, 지시사항에 대해 주의 깊게 고려한다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**16. I brainstorm and write down ideas before I begin to write in English. \***

나는 영어로 글을 쓰기 전 브레인스토밍하고 떠오르는 것들을 적는다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**17. I make plans and outlines in Korean before I begin to write in English\***

나는 영어로 글을 쓰기 전 한국어로 계획을 세우거나 개요를 짠다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**18. I make plans and outlines in English before I begin to write in English. \***

나는 영어로 글을 쓰기 전 영어로 계획을 세우거나 개요를 짠다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**19. I make plans and outlines both in Korean and English before I begin to write in English. \***

나는 영어로 글을 쓰기 전 한국어와 영어로 계획을 세우거나 개요를 짠다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**20. I make a timetable before I begin to write in English. \***

나는 영어로 글을 쓰기 전 시간 계획을 짠다.

1 2 3 4 5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**21. I notice vocabulary related to a topic that I will write about and try to remember the words.\***

나는 내가 써야할 주제와 관련된 어휘들을 생각하고 그 단어들을 기억하려고 노력한다.

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

## Section D (22~28)

**22. I like to write Korean first and then translate it into English.\***

나는 한국어로 먼저 글을 쓴 후 영어로 번역을 한다.

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

**23. I usually keep my plan as I am writing.\***

나는 글을 쓰면서 글을 쓰기 전 세웠던 계획을 지킨다.

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

**24. I like to change or make my ideas clearer as I am writing.\***

나는 글을 쓰면서 나의 생각을 바꾸거나 더 분명히 다듬는다.

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

**25. I make up new words if I do not know the right ones when I am writing in English.\***

나는 글을 쓰면서 적절한 영어 단어를 모를 때는 새로운 단어를 만들어낸다.

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

**26. If I can't think of a specific English word, I use a different word or phrase that means the same thing.\***

나는 특정 영어단어가 생각나지 않으면 같은 의미를 지닌 다른 단어나 어구를 사용한다.

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

**27. I edit for content and ideas as I am writing. \***

나는 글을 쓰면서 내용과 생각들을 편집하고 재정비한다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

**28. I edit for organization as I am writing.**

나는 글을 쓰면서 글의 구성을 편집하고 재정비한다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

## Appendix 3

### Diagnostic Test for Writers

Read the directions carefully before beginning each section. There are **25** questions in the test. You will have a total of **15 minutes**. Do not spend too much time on any single question. It is a diagnostic test to help you identify your strengths and weaknesses in English skills.

#### Section 1

Sample: Each of the questions below contains an error in one of the underlined parts. Choose the correct answer that corresponds to the part of the sentence containing the error.

*Example:*

When you recieve the prize, your family will celebrate the happy occasion.

A

B

C

D

Answer: A is the right answer because the spelling of the word should be corrected to 'receive'.

#### Section 2

Sample: In the following questions, part of each item is underlined. It may or may not be correct. Immediately following are four ways of writing the underlined part. Choose the best one. Choice A repeats what is underlined; choose it if the original needs no revision.

*Example:*

Hunting is limited in most states. People with special licenses are only allowed to hunt.

A. People with special licenses are only allowed to hunt.

B. Only people with special licenses are allowed to hunt.

C. People with special licenses are allowed only to hunt.

D. People with special licenses are allowed to hunt only.

Answer: B is the correct answer.



### Section 3

Sample: Following are thesis statements, each with three supporting topic sentences.

Fill in a topic sentence and a thesis statement.

*Example:*

1. *Thesis Statement:*

This village is the best environment for me to live in.

*Topic sentences:*

1. I like its excitement.
2. I like the availability of resources.
3. \_\_\_\_\_

**Possible answer:** I like its security and friendly people.

 Look at the following picture. Describe what you see in English.

*Example:*



**Possible answer:** It shows a father and son on a fishing trip. The father has got a tiny fish and his son has a giant fish.

**Now, let's begin the actual test.**

## Section 1

☞ Each of the questions below contains an error in one of the underlined parts. Choose the correct answer that corresponds to the part of the sentence containing the error. (1~11)

1. The Parkers told my neighbors that their leaving for Kansas City in the morning.  
A B C D
2. During the Middle Ages, scribes copved books carefully by hand using quill pens.  
A B C D
3. The attorneys new office was similar to her former one, except that it was on a quieter street.  
A B C D
4. Often the reputation of an entire company depend on one employee who officially represents that company to the public.  
A B C D
5. Either two semesters of a foreign language or a passing grade on a moderately difficult translation examination fulfill the college's foreign language requirements for students who plan to major in engineering, including you and me.  
A B C D
6. Mother advised us girls not to lay in the sun because the temperature was extremely high.  
A B C D
7. When Ben realized that he had set on the train longer than an hour, he quickly asked the gentleman beside him for the correct time.  
A B C D
8. All the candidates for student office, including you and her, have filed the necessary papers with the university.  
A B C D



- C. a college freshman. She has returned to school because she wants to become a lawyer.
- D. a college freshman; she has returned to school. Because she wants to become a lawyer.

14. Environmentalists are alarmed about coal mining. Which has left ugly scars on the land.

- A. Environmentalists are alarmed about coal mining. Which has left ugly scars on the land.
- B. Environmentalists are alarmed. About coal mining. Which has left ugly scars on the land.
- C. About coal mining which has left ugly scars on the land, environmentalists are alarmed.
- D. Environmentalists are alarmed about coal mining, which has left ugly scars on the land.

15. Tuition is going up next year. Because the college's expenses have risen. Fuel, maintenance, and insurance costs are all higher than last year.

- A. Tuition is going up next year. Because the college's expenses have risen. Fuel, maintenance, and insurance costs are all higher than last year.
- B. Tuition is going up next year because the college's expenses have risen, fuel, maintenance, and insurance costs are all higher than last year.
- C. Tuition is going up next year. Because the college's expenses have risen, fuel, maintenance, and insurance costs are all higher than last year.
- D. Tuition is going up next year because the college's expenses have risen. Fuel, maintenance, and insurance costs are all higher than last year.

16. My college offers many interesting courses to first year students. Therefore, I do not know. What electives to take.

- A. My college offers many interesting courses to first year students. Therefore, I do not know. What electives to take.
- B. My college offers many interesting courses. To first year students. Therefore, I do not know. What electives to take.

- C. My college offers many interesting courses to first year students. Therefore, I do not know what electives to take.
- D. My college offers many interesting courses to first year students therefore, I do not know what electives to take.

17. Many people own Turkish rugs who have never been there.

- A. Many people own Turkish rugs who have never been there.
- B. Many people own Turkish rugs but have never been there.
- C. Many people who own Turkish rugs have never been to Turkey.
- D. Many people owning Turkish rugs have never been there.

18. Flying over Paris, many lights could be seen by the passengers.

- A. Flying over Paris, many lights could be seen by the passengers.
- B. Many lights could be seen by the passengers flying over France.
- C. Flying over Paris, the passengers saw many lights.
- D. The passengers saw many lights flying over Paris.

19. Managing a nuclear power plant requires great attention to detail. You have to be careful to follow procedures so that no radioactivity escapes.

- A. You have to be careful to follow procedures so that no radioactivity escapes.
- B. Procedures you should follow should not let any radioactivity escape.
- C. You have to be very careful to follow procedures. Don't let any radioactivity escape.
- D. Procedures must be followed exactly to guarantee that no radioactivity escapes.

20. The doctor operated swiftly, describing the procedures as she went along.

- A. The doctor operated swiftly, describing the procedures as she went along.
- B. The doctor operated swiftly and describing the procedures as she went along.
- C. Operated swiftly, the doctor described the procedures as she went along.
- D. Operated swiftly, the procedures as she went along the doctor described.

### Section 3

✎ Following are thesis statements, each with three supporting topic sentences. Fill in a topic sentence and a thesis statement. (21~22)

21. *Thesis Statement:*

Smoking cigarettes is harmful to your health.

*Topic sentences:*

2. Heavy cigarette smoking causes throat diseases.

3. Smoking can damage the lungs.

4. \_\_\_\_\_

22. *Thesis Statement:*

Watching television is \_\_\_\_\_

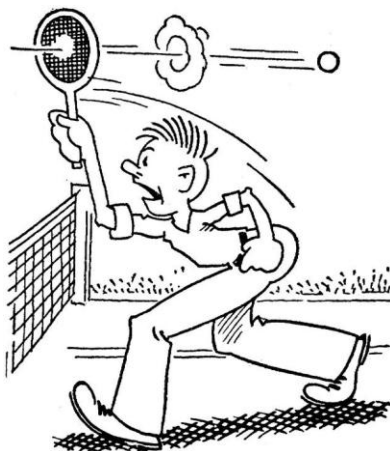
*Topic Sentences:*

1. It is a valuable educational tool.

2. It helps us to relax.

3. It provides something for our family to discuss.

✎ Look at the following pictures. Describe what you see in English. (23~25)



23.



24.



25.

**This is the end of the test. You may check your answers again.**

## Appendix 4

### Writing Prompts

#### *Pretest Directions for every group*

☞ You will be asked to write an essay in which you state, explain, and support your opinion on an issue. You will have 30 minutes to complete your essay.

#### *Posttest Directions*

☞ You will be asked to write an essay in which you state, explain, and support your opinion on an issue.

#### *For group 2*

You will have 5 minutes to plan before you write. You can take notes on the sheet provided. Then, you have 25 minutes to write your essay.

#### *For group 3*

You will have 30 minutes to plan, write, and revise your essay. You can take notes on the sheet provided.

#### *Prompt A*

In some countries, people are no longer allowed to smoke in many public places and office buildings. Do you think it is a good rule or a bad rule?

#### *Prompt B*

People have different ways of escaping the stress and difficulties of modern life. Some read: some exercise: others work in their gardens. What do you think the best ways of reducing stress?



## Appendix 5

### Post-questionnaire

Please read each statement below, then fill in the circle under the number that most accurately reflects your opinion/attitude, or write an answer.

다음에 나오는 각각의 질문에 답하시거나, 의견에 부합하는 숫자나 상자에 체크해 주세요.

\*

#### 1. The time allowed for the task was too short. \*

시험시간이 너무 짧았다고 생각한다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

#### 2. I felt I needed more planning time. \*

준비시간이 더 필요하다고 느꼈다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------

#### 3. How much time do you think is appropriate for planning before writing? \*

준비 시간은 얼마 정도가 적당하다고 생각하나요?

- ☐ 1 to 2 minutes
- ☐ 3 to 5 minutes
- ☐ 5 to 7 minutes
- ☐ 7 to 10 minutes
- ☐ more than 10 minutes

#### 4. I felt the task was too difficult. \*

시험이 너무 어려웠다.

1      2      3      4      5

Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
-------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	----------------






#### 5. Which topic did you feel was more difficult? \*

두 주제 중 어느 주제가 더 어려웠습니까?

- ☐ Banning smoking in public areas 공공장소에서의 흡연 금지
- ☐ Ways to relieve stress 스트레스 해소 방법
- ☐ Neither of them was difficult. 두 주제 모두 어렵지 않았다

**6. I believe that I was able to show my English writing ability through the task.\***

이 시험을 통해 나의 영어글쓰기 능력을 보여줄 수 있었다고 생각한다.

	1	2	3	4	5	
Strongly disagree						Strongly agree

**7. Have you ever learned how to plan before writing in English?\***

영어글쓰기 전 어떻게 준비하고 계획하는지에 대해 학습한 경험이 있나요?

- ☐ Yes
- ☐ No

- ☒ ☐ No

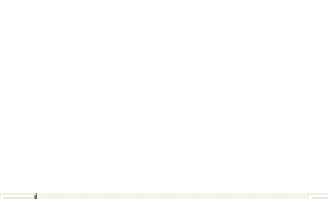
8. If you said 'yes' for the previous question, how did you practice making plans? Please explain in detail.\*

학습경험이 있다면, 어떻게 개요짜기를 연습했나요?



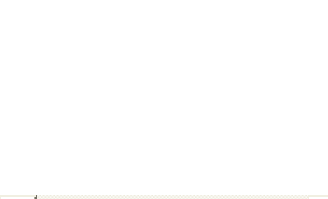
9. What do you usually do when you make plans before writing? Please explain in detail.\*

글쓰기 전 준비할 때 주로 어떤 계획을 세우나요? 자세히 설명해주세요.



10. Do you think it is necessary to make plans before you write? Please explain why or why not.\*

영어 글쓰기를 하기 전에 계획을 세울 필요가 있다고 생각하나요? 설명해주세요.



## Appendix 6

### Analytic Scoring Rubrics for TOEFL CBT Writing Prompts

#### Scoring Rubrics for Development of Ideas

##### ***Level 1: Low English Proficiency***

Due to problems with English Proficiency, the main points are very difficult to understand. The words may not be in the form of standard English (confusing word order, wrong word forms, frequent misspellings), so that it is hard to know what the writer is trying to say.

##### ***Level 2: Limited Response***

Due to the limited response given, the development of ideas cannot be judged. Because the response has fewer than eight full typed lines of text (or fewer than 90 words), there is not enough evidence to judge development

##### ***Level 3: Minimal Development***

Only a few of the main points (less than half) are developed with supporting details, explanations or brief examples. One of the reasons may be developed with a brief example, but the other reasons are simply stated. Or, an explanation is given for part of the writer's opinion, but the rest of the opinion is not developed. Or the essay is mostly a list of ideas supporting the opinion or discussing the theme.

##### ***Level 4: Basic Development***

Most of the main points (half or more) are developed to one or two levels of depth, although some of the main points may not be developed. One common pattern is: reason → explanation; reason → explanation; reason → explanation. Another is: main point → explanation → example; main point → example; main point → explanation; main point.

##### ***Level 5: Some Depth of Development***

One of the main ideas is developed in depth, to the third level of development, such as:

main ideas → explanation → example → conclusion or main point → problem → solution → example. All of the ideas may not be developed or may only be somewhat developed, but one main idea is developed in depth.

### ***Level 6: More Depth of Development***

At least two of the main points are developed in depth (to the third level). For the academic debate essay, a common pattern is: reasons for opinion → explanation → example → conclusion. For the kids and sports essay, a common pattern is idea → explanation → problem → solution.

## **Scoring Rubric for Organization**

### ***Level 1: Low English Proficiency***

Due to problems with English proficiency, the points the writer is trying to make are unclear. The words may not be in the form of standard English and/or the words may be so out of order that one cannot understand the essay.

### ***Level 2: Limited Response***

Due to the limited response given, the organization of ideas cannot be judged. Because the response has less than eight typed lines of text (or few than 90 words), there is not enough evidence to judge organization

### ***Level 3: Some Organization of Ideas***

Some of the ideas flow logically, but most read more like a list of ideas about the topic. In one or two parts of the essay, the writer made some decisions about how to present ideas, i.e., how to order ideas to make a point. However, many of the ideas do not flow logically, the writer changes direction suddenly, interrupting the flow, often making it hard for the reader to understand the main points.

### ***Level 4: More Organization of Ideas***

Most of the ideas flow logically (although there still may be a few sudden changes in direction). The writer has clearly made decisions about how to order ideas to make a point, making it easier for the reader to understand the main points.

### ***Level 5: Basic Overall Essay Structure***

There is an overall structure to the essay, but it is very basic. The writer may use the prompt to structure the essay (i.e., discussing advantages in one paragraph and disadvantages in another or discussing reasons to support an opinion). OR the structure provided by the thesis statement is not followed. Within the essay, most of the ideas flow logically (although there may be a few sudden changes in direction)

### ***Level 6: Advanced Overall Essay Structure***

The overall structure to the essay is very clear and involves an

organizing principle or theme that goes beyond the structure of the prompt. Those using the Road Map approach, articulate their organizational structure at the beginning of their essays. Those using the Journey of Discovery approach, articulate their organizational structure near the end of their essays. Within the essay, most of the ideas flow logically (although there may be sudden changes in direction).

### **Scoring Rubric for Vocabulary**

#### ***Level 1: Not Enough Evidence***

Due to the limited response given, the writer's command of vocabulary cannot be judged.

Because the response has less than eight full typed lines of text (or less than 90 words), there is not enough evidence to judge vocabulary.

#### ***Level 2: Basic***

The essay is mostly comprised of basic words. The range of words is limited to simple expressions, words copied from the prompt, and basic vocabulary that is often used repeatedly. Papers that are longer than eight full lines of text but are difficult to understand are also classified as Basic.

#### ***Level 3: Predictable***

The essay now includes a mixture of descriptive words and basic words. But most of these words are within a predictable range for students at this level (those taking the TOEFL exam).

#### ***Level 4: More Varied***

More of the words are descriptive and a wider range of these words is now used. There may be an attempt use more specialized words, but these words are not used correctly.

#### ***Level 5: Effective***

At this level, all three types of words are used including specialized words, and the range of vocabulary is effective. This represents a more sophisticated control over vocabulary.

\* Please note that words provided in the prompt do not count towards the vocabulary rating

\* As students begin to use more sophisticated words, they often misspell these new words. The misspellings count as mechanics errors, and the word contributes towards the range of vocabulary.

\* However, if they misuse a word, this word does not contribute towards the range of vocabulary.

## **Scoring Rubric for Sentence Variety and Construction**

### ***Level 1: Not Enough Evidence***

Due to the limited response given, the writers' command of sentence variety and construction cannot be judged. Because the response has less than eight full typed lines of text (or less than 90 words), there is not enough evidence to judge variety and construction.

### ***Level 2: Minimal Control***

Mostly simple sentence structures are used, with little variety; OR almost all of the sentences have the same structure; OR the order of words is so irregular that it is hard to understand the main points.

### ***Level 3: Some Control***

There is some variety in sentence structure; some more complex structures are used.

However, the attempt to use more complex structures often results in awkwardly constructed sentences.

### ***Level 4: Adequate Control***

A wider variety of sentence structures is used. While some of the complex sentences may be awkward, others are well structured.

### ***Level 5: Basic Overall Essay Structure***

Writers use a variety of sentence structures to effectively convey the main points. Most of the more complex sentences are well-structured.

## **Scoring Rubric for Grammar and Usage**

### ***Level 1: Not Enough Evidence***

Due to the limited response given, the writer's pattern of grammatical errors cannot be judged. Because the response has less than eight full typed lines of text (or less than 90 words), there is not enough evidence to judge the writer's control over usage.

### ***Level 2: Minimal Control***

Grammatical errors are constant—75% of the sentences have grammatical errors. OR the grammatical errors are so serious that it is hard to understand the main points.

### ***Level 3: Some Control***

There are frequent errors across the paper, but the errors do not interfere with understanding the main points. More than half of the sentences contain grammatical errors (51–74%).

***Level 4: Adequate Control***

There are not as many grammatical errors across the paper and these errors do not interfere with understanding the main points and subpoints. Half or less than half of the sentences contain grammatical errors (26–50%). Also, the types of errors tend to be aspects of usage that are acquired at later stages of second language development (such as the rules for the use of prepositions and articles).

***Level 5: Strong Control***

There are few, minor grammatical errors across the paper so that it is easy to understand the main point and subpoints. One-quarter or less than one-quarter of the sentences contain grammatical errors (0–25%). The types of errors are aspects of usage that are acquired at later stages of second language development.

**Scoring Rubric for Mechanics*****Level 1: Not Enough Evidence***

Due to the limited response given, the writer's command of mechanics cannot be judged. Because the response has less than eight full typed lines of text (or less than 90 words), there is not enough evidence to judge mechanics.

***Level 2: Minimal Control***

Mechanical errors are constant—75% of the sentences have mechanical errors. OR the errors in mechanics are so serious that it is hard to understand the main points. This sometimes happens with frequent spelling and punctuation errors.

***Level 3: Some Control***

There are frequent errors across the paper, but the errors do not interfere with understanding the main points. More than half of the sentences contain errors in mechanics (51–74%).

***Level 4: Adequate Control***

There are not as many errors across the paper and the errors do not interfere with the understanding of the main points and most of the subpoints. Half or less than half of the sentences contain errors in mechanics (26–50%).

***Level 5: Strong Control***

There are few errors across the paper. One-quarter or less than one-quarter of the sentences contain errors in mechanics (0–25%)

## **Appendix 7**

### **Post-rating Questionnaire**

1. What was your overall opinion about the writing samples?
  
  
  
  
  
  
  
  
  
  
2. What was the most difficult part when rating the writing samples?
  
  
  
  
  
  
  
  
  
  
3. How would you rank the analytic criteria according to their rating difficulty?  
(0 being the least difficult and 5 being the most difficult).

\_\_\_\_\_ Development of Ideas  
\_\_\_\_\_ Organization  
\_\_\_\_\_ Vocabulary  
\_\_\_\_\_ Sentence Variety and Construction  
\_\_\_\_\_ Grammar and Usage  
\_\_\_\_\_ Mechanics

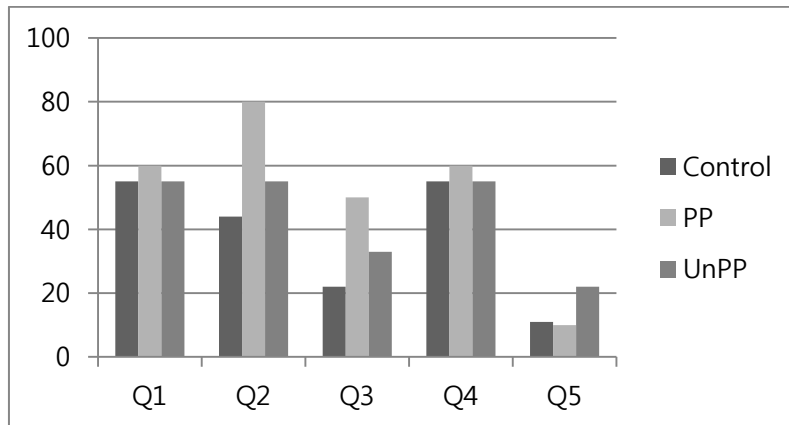
Why did you choose the particular order?

4. Do you have any other comments that are not addressed in the previous questions?



## Appendix 8

### Results of Self assessment of English Writing Strategies (in part)



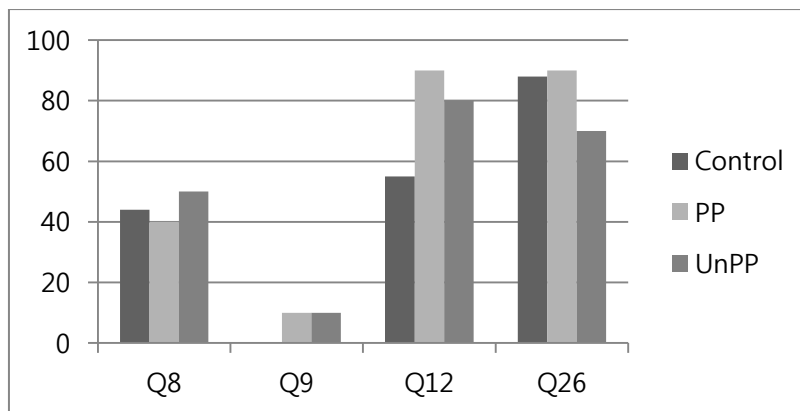
Q1. I try to find as many as I can to use English.

Q2. I look for opportunities to read as much as possible in English.

Q3. I look for opportunities to speak as much as possible in English

Q4. I look for opportunities to listen as much as possible in English.

Q5. I look for opportunities to write as much as possible in English

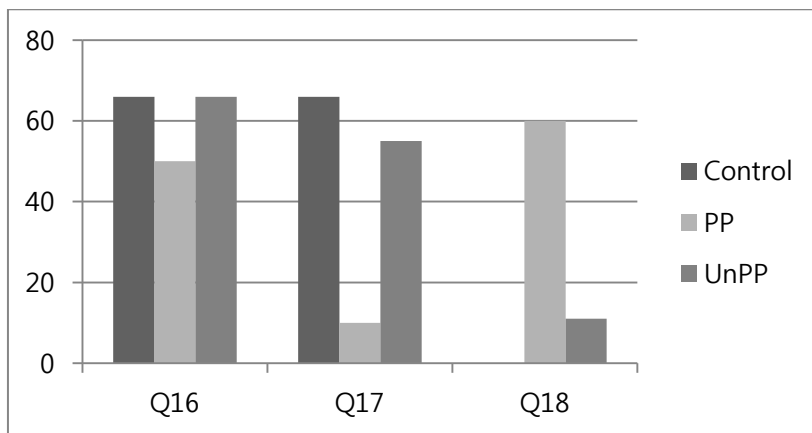


Q8. I often write in Korean.

Q9. I often write in English.

Q12. I use cohesive devices in my English writing (e.g. thus, moreover, however and so on).

Q26. If I can't think of a specific English word, I use a different word and phrase that means the same.



- Q16. I brainstorm and write down ideas before I begin to write in English.  
Q17. I make plans and outline in Korean before I begin to write in English.  
Q18. I make plans and outlines in English before I begin to write in English.

## 국문초록

# 과제 내 계획하기 조건들이 한국인 EFL 수험자들의 논증적 글쓰기 과제 수행에 미치는 영향

서울대학교 대학원  
영어영문학과  
박 정 현

글쓰기에 있어서 ‘계획하기(planning)’는 학습자들이 자신의 생각을 정리하고 조직하고 목표를 세우는 인지 과정에서 매우 중요한 단계로서 인식되어 왔다. 이에 따라 다양한 연구들이 ‘계획하기’의 효과를 밝히기 위해 이루어졌고, 상충되는 결과들을 보고해왔다. 쓰기 언어 평가의 측면에서는 쓰기 과업뿐만 아니라 쓰기과업을 수행하는 조건, 맥락(task conditions)에 대한 논의가 지속적으로 이루어져 왔다. 그러나 상대적으로 쓰기 언어 평가의 상황에서 ‘계획하기’ 조건(planning conditions)에 대한 연구는 거의 이루어지지 않았다.

본 연구의 목적은 한국인 EFL 학습자들의 논증적 글쓰기 평가과제(argumentative writing assessment tasks)에서의 ‘과제 내 계획하기’ 조건의 효과에 대해 밝히려는 것이다. 과제 내 계획하기(within-task planning)는 수험자가 쓰기 평가 과제를 수행하는 중에 계획하기의 시간적 제한이 주어지는 경우 (pressured)와 시간적 제한은 주어지지 않으나 계획하기를 쓰기 과정에 포함시켜 유도하는 경우가(unpressured) 있다. 상급, 중급의 영어능력을 지닌 28명의 한국인 대학생이 세 가지 환경(계획하기 과정이 없는 통제집단, 계획하기 시간 제한이 주어진 실험 집단, 시간 제한은 없으나 계획하기 과정이 포함된 실험 집단)에 배정되어 두 번의 논증적 글쓰기 쓰기 과제를 수행하였다. 수험자가 수행한 쓰기 과제는 주제의 전개, 조직, 어휘, 문장 구성, 문법과 철자 및 구두법의 총 여섯 가지 분석적 영역으로 평가되었다.

‘계획하기’ 조건의 효과는 분석적 영역 점수와 그 합산 점수를 바탕으로 반복측정 분산분석(repeated ANOVA)을 통해 측정되었고, 수험자의 영어능력이 계획하기 조건의 효과를 중재시키는 요인이었는지를 알아보기 위해 공변량 분석(ANCOVA)을 실시하였다. 또한, 한국인 EFL 수험자들의 계획하기 행동과 패턴, 쓰기 전략을 알아보기 위해 실험 전 후 설문지와 인터뷰를 실시하여 살펴보았다.

분석 결과 과제 내 계획하기 조건은 수험자의 쓰기 과제에 있어서 각 분석적 영역 점수와 합산 점수에 유의미한 영향을 미치지 않음을 보여주었다. 그리고 수험자의 영어 능력도 분석적 점수뿐만 아니라 합산 점수에 통계적으로 유의미한 영향을 미치지 않았다. 반면, 수험자의 75%가 계획하기를 쓰기 과제를 수행하는데 있어서 중요한 개념을 정리하는 것에 도움이 되는 핵심적인 과정으로 인식하고 직접 적용하고 있음이 드러났다. 영어 능력 상급 수험자가 중급 수험자보다 계획 하기에 더 많은 시간을 소모하는 것도 나타났다.

본 연구는 과제 내 계획하기 조건과 시간을 할당하여 쓰기 과제 평가를 구성해야 하는 가에 대해 함의하고 있다. 요컨대, 과제 내 계획하기 조건은 수험자의 쓰기 과제 수행 점수에 통계적으로 유의미한 영향을 미치지 않으므로 현재 쓰기 과제 평가의 구성 방식, 즉 과제 내에 계획하기 조건과 시간을 따로 제한하지 않는 형식에 대해 경험적 근거를 제시하고 있다.

**주요어 :** 과제 내 계획하기 조건, 계획하기 조건의 영향, 논증적 글쓰기 평가 과제, 수험자 언어능력, 계획하기 패턴

**학번:** 2009-22753